EXTERNAL SCIENTIFIC REPORT

Investigation of the state of the art on identification of appropriate reference points for the derivation of health-based guidance values (ADI, AOEL and AAOEL) for pesticides and on the derivation of uncertainty factors to be used in human risk assessment[1]

Chemicals Regulation Directorate, Health & Safety Executive, UK

ABSTRACT

This project evaluated current approaches and alternative methodologies to the derivation of health based guidance values for chemical exposures. Although the report relies on evaluations of pesticides the basic considerations apply to other areas of chemical risk assessment. Pesticide evaluations were reviewed to determine studies and endpoints utilised to derive guidance values, safety factors applied and any aspects routinely debated during peer review. Approaches that would be applicable to the derivation of the recently proposed 'Acute Acceptable Operator Exposure Level' and supporting data were evaluated. A common topic of discussion during the interpretation of toxicological data is whether effects seen in animal studies are relevant to humans. Literature relating to end-points routinely used in deriving reference values was evaluated together with responses to a questionnaire and proposals developed for producing a more consistent approach. Alternative approaches to the No Observed Adverse Effect Level (NOAEL) were evaluated. Particular consideration was given to the Benchmark Dose (BMD) approach, with a number of case-studies performed to determine practicalities of current software programs. Alternative approaches to the use of the default 100 fold safety factor to address uncertainties in extrapolating between animal data and human exposures were evaluated. Conclusions and recommendations included: Current approaches are protective of human exposures but there is potential for improvement using alternative methods and revised test guidelines. The ARfD can be a basis for derivation of AAOELs. BMD offers significant benefits and should be utilised routinely on the end-points used to derive guidance values. A number of findings in animal studies might be of no relevance to humans but this needs to be demonstrated in each case. Allometric scaling is a viable alternative approach whereas CSAF and PBPK are too data intensive for general use. Description of the uncertainty surrounding guidance values should be improved.

KEY WORDS

Health based guidance value; Pesticides; Chemicals; Benchmark dose; adaptive response; uncertainty; risk assessment.

DISCLAIMER

The present document has been produced and adopted by the bodies identified above as author(s). This task has been carried out exclusively by the author(s) in the context of a contract between the European Food Safety Authority and the author(s), awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety

[1] Question No EFSA-Q-2011-00322

Any enquiries related to this output should be addressed to pesticides.ppr@efsa.europa.eu

Suggested citation: Chemicals Regulation Directorate, Health & Safety Executive, UK; Investigation of the state of the art on identification of appropriate reference points for the derivation of health-based guidance values (ADI, AOEL and AAOEL) for pesticides and on the derivation of uncertainty factors to be used in human risk assessment. Supporting Publications 2013:EN-413. [169 pp.]. Available online: www.efsa.europa.eu/publications

Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

TABLE OF CONTENTS

BACKGROUND AND TERMS OF REFERENCE AS PROVIDED BY THE REQUESTOR

Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC, stipulates in Article 4 (1) and Annex II paragraph 3.1. that for the approval of an active substance, where relevant, acceptable daily intake (ADI) and acceptable operator exposure level (AOEL) shall be established.

The ADI of a chemical is the estimate of the amount of a substance in food or drinking water, expressed on a body weight basis, than can be ingested daily over a lifetime without appreciable health risks to the consumer on the basis of all known facts at the time of the evaluation (WHO, 1997).

Currently, there is no harmonised and internationally agreed guidance for setting the acceptable daily intake of the pesticide active substances to assist the European Commission and Member States when making decisions about inclusion of an active substance in Annex I of Directive 91/414/EEC.

The ADI concept was introduced in 1957 by the Council of Europe and later on was taken over by the Joint FAO/WHO Meeting on Pesticide Residues (JMPR). To date the ADI is considered as being a valid tool in risk assessment (Galli et al., 2008). Traditionally ADI values are based on reference points (RPs) derived from results of long-term animal toxicity studies. An uncertainty factor is usually applied to convert the relevant reference point, normally No Observed Adverse Effect Levels (NOAELs) or Lowest Observed Adverse Effect Levels (LOAELs), in the absence of relevant NOAELs, into a safe daily intake value for humans. In recent publications the appropriateness of this approach has been questioned since setting of NOAELs/LOAELs is dependent on study design, dose selection, group size and the precision with which the test is performed (Renwick et al., 2003). The Benchmark Dose (BMD) approach arose as an alternative way of defining reference points for risk assessment. It was concluded in a recent opinion of EFSA‟s Scientific Committee, that the BMD approach is a scientifically more advanced method as compared to the NOAEL approach for deriving a reference point, since it also includes consideration of the dose-response curve and quantifies uncertainty and variability of dose-response data. (EFSA, 2009).

Historically, the currently routinely applied uncertainty factor (UF) of 100 was introduced in 1954 by Lehman and Fitzhugh. It is designed to reflect application of two separate 10-fold factors for inter-species differences (animal-human) and human variability respectively (WHO, 1987).

The traditional inter-species factor of 10 was later divided into values of 4 for differences in kinetics and 2.5 for toxicodynamic differences (Renwick et al., 1993).

The subdivision of the UF for intra-species variation evenly to 3.16-fold for both kinetics and dynamics (IPCS, 1994) permits the use of specific data on a chemical to derive chemical-specific adjustment factors (CSAF). Compound specific data for one particular aspect of uncertainty should be used to replace the relevant part of the overall default uncertainty factor (WHO, 2001: 2005).

In recent publications human variability in kinetics has been analysed for the main metabolic pathways in healthy adults and subpopulations of the elderly, neonates and children. Based on this analysis pathway-related UFs could be established that allow the incorporation of metabolism data into the derivation of health-based guidance values. (Dorne et al., 2004). In a recent review four main scenarios were identified in humans for which the current default uncertainty factor for toxicokinetics (3.16) does not cover the human variability (Dorne, 2010). The appropriateness of the current

approaches to deal with the uncertainty in the establishment of health-based guidance values should therefore be evaluated and, if relevant, other science-based approaches of application and refinement of uncertainty factors in setting the ADI should be elaborated.

According to Directive 97/57/EC (establishing Annex VI to Directive 91/414/EEC), the acceptable operator exposure level (AOEL) is the maximum amount of active substance, expressed on a body weight basis, to which the operator may be exposed without any adverse health effects. Thus the AOEL is a health-based exposure limit to be used for comparison with estimated or measured exposure levels of operators, workers and bystanders for assessing the risk of these groups arising from the application of a plant protection product.

AOELs should normally be based on reference points (i.e. NOAELs, alternatively in some cases also LOAELs) that are obtained in short-term animal toxicity studies. Based on assumptions similar to those described for setting the ADI, an uncertainty factor of 100 is also applied to derive an AOEL (EC, 2006). However, very similar considerations and criticisms, as described above for setting ADIs, are valid for the appropriateness of the way of setting the AOEL values (both for deriving reference points and uncertainty factors).

Moreover, EFSA's Panel on Plant Protection Products and their Residues (PPR) recommended, in a recently adopted opinion on Preparation of a Guidance Document on Pesticide Exposure Assessment for Workers, Operators, Bystanders and Residents, that guidance should be developed on the derivation of a new reference value, a so called "acute AOEL" or "AAOEL", that is required for an acute risk assessment for operators, workers and bystanders for PPPs that are acutely toxic (EFSA, 2010). Therefore the existing Draft Guidance for the Setting and Application of Acceptable Operator Exposure Levels (AOELs) (SANCO 7531 - rev.10, 2006) should be updated to provide guidance on how to derive and apply this new health-based guidance value.

In addition to the above, the Pesticide Steering Committee (an EFSA Network providing a platform for cooperation and consultation between the different actors involved in pesticide risk assessment in the EU), very recently identified the necessity of giving high priority to the development of a new guidance on setting the acceptable daily intake of pesticide active substances and the revision of the existing draft guidance for setting and application of acceptable operator levels..

REFERENCES:

Dorne, J.L.C.M, 2010. Metabolism, variability and risk assessment. Toxicology, 268, 156-164.

Dorne J.L.C.M, Walton K, Renwick A.G, 2004. Human variability in xenobiotic metabolism and pathway-related uncertainty factors for chemical risk assessment: a review. Food and Chemical Toxicology 43, 203–216.

EC (European Commission), 2006. Draft Guidance for the Setting and Application of Acceptable Operator Exposure Levels (AOELs) (SANCO 7531 - rev.10, 2006).

EFSA (European Food Safety Authority), 2009. Guidance of the Scientific Committee on a request from EFSA on the use of the benchmark dose approach in risk assessment. The EFSA Journal (2009) 1150, 1-72

EFSA Panel on Plant Protection Products and their Residues (PPR); 2010. Scientific Opinion on Preparation of a Guidance Document on Pesticide Exposure Assessment for Workers, Operators, Bystanders and Residents..2010;8(2):1501, 24 pp.

Galli C.L, Marinovicha M, Lotti M, 2008. Is the acceptable daily intake as presently used an axiom or a dogma? Toxicology Letters, 180, 93–99.

Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC, OJ L 309, 24.11.2009, p. 1-50

Renwick A.G, Barlow S.M, Hertz- Picciotto I, Boobis A.R, Dybing E, Edler L, Eisenbrand G, Greig JB, Kleiner J, Lambe J, Müller D.J.G, Smith M.R, Tritscher A, Tuijtelaars S, van den Brandt P.A, Walker R, Kroes R, 2003. Risk Characterisation of chemicals in food and diet. Food and Chemical Toxicology, 41, 1211-1271.

WHO (World Health Organization), 1987. International program on Chemical Safety:: Principles for the safety assessment of food additives and contaminants in food, Environmental Health Criteria, 70, 174pp,

WHO (World Health Organization), 1994. International Programme on Chemical Safety: assessing human health risks of chemicals: derivation of guidance values for health-based exposure limits. Environmental Health Criteria, 170, 73pp,

WHO (World Health Organization), 1997. Guidelines for predicting dietary intake of pesticide residues, 34 pp.

WHO (World Health Organization), 2001. International Programme on Chemical Safety: guidance document for the use of chemical-specific adjustment factors (CSAFs) for interspecies differences and human variability in dose–concentration response assessment. 76pp,

WHO (World Health Organization), 2005. International Programme on Chemical Safety: Chemical-specific adjustment factors for Interspecies differences and human variability: Guidance document for use of data in dose/concentration–response assessment. IPCS harmonization project document; no. 2. ISBN 92 4 154678 6

PURPOSE OF THE ASSIGNMENT

Overall objective:

The overall objective of the contract resulting from the present procurement procedure is to investigate the state of the art on the identification of the appropriate Reference Point for the derivation of health-based guidance values for pesticides (ADI, AOEL and AAOEL) and on the derivation of respective uncertainty factors in view of developing EU guidance on the setting of the ADI, AOEL or AAOEL of active substances of pesticides.

Specific objectives:

The specific objectives of the contract resulting from the present procurement procedure are as follows:

- Information collection related to the principles of identifying and characterising the critical reference points for setting the health-based guidance values of chemicals in or on food.

- Information gathering and elaboration of general criteria to evaluate the appropriateness of toxicological effects to be used for derivation of reference points for setting health-based guidance values.

- Evaluate the appropriateness of the current approaches of the derivation of uncertainty factors for converting identified reference points into health-based guidance values.

- Propose a science based approach of application and refinement of the uncertainty factors to be applied in setting the ADI, AOEL and AAOEL values.

SCOPE OF THE WORK, EXPECTED OUTCOMES AND DELIVERABLES, TIMELINE AND PAYMENTS

The contractor selected as a result of this tendering procedure is expected to carry out the following tasks:

- systematically and comprehensively evaluate EFSA"s Conclusions on ADI and AOEL setting with an emphasis on the rationales used and generic expert discussions for the selection of critical study, derivation of the relevant reference points (i.e. NOAELs and LOAELs), and the derivation of uncertainty factors;

- systematically and comprehensively evaluate EFSA"s Conclusions on the ARfD (Acute Reference Dose) setting similarly to ADI and AOEL evaluations with an emphasis on exploring and gauging the usefulness of these evaluations in regard to the development of a concept for deriving a possible future "AAOEL";

- comprehensively collect and systematically evaluate relevant scientific literature or other relevant publications or information sources (like guidance documents or government reports) related to the principles of identifying and characterising the critical reference points for hazard and risk assessments outside the EU and outside pesticide evaluations that could be relevant for future ADI, AOEL and "AAOEL" settings for pesticides in the EU (e.g. derivation of TDIs);

- collect and evaluate possible alternative reference points (to NOAEL and LOAEL) to be used for the derivation of AOELs, AAOELs and ADIs with a particular emphasis on the evaluation of the BMD approach and duly taking into account relevant EFSA publications;

- collect and evaluate relevant scientific literature or other relevant publications or information sources (e.g. guidance documents, government reports) related to the understanding or assessment of the adverse or adaptive nature of effects observed in toxicological studies and provide criteria to evaluate the appropriateness of toxicological effects to be used for derivation of reference points for setting health-based guidance values;

- collect and scrutinize relevant scientific literature or other relevant publications or information sources (e.g. guidance documents or government reports) on the appropriateness of the current approaches to deal with the uncertainty in the establishment of health-based guidance values and, if relevant, propose other science-based approaches of application and refinement of uncertainty factors in setting the ADI, AOEL or "AAOEL" values.

The contractor is expected to work closely with EFSA throughout the course of the project. EFSA will set up a project steering group to accompany this project and the contractor will be expected to cooperate with this group.

The following coordination meetings with EFSA are foreseen:

1. Kick off meeting: The purpose is to discuss the objectives and scope of the project in detail. It should take place within 4 weeks after the signature of the service contract resulting from the present procurement procedure.

2. Interim meeting by month 6: The purpose is to clarify outstanding questions and to discuss the interim report.

3. Final meeting at the beginning of month 12: The purpose is to discuss the draft final report.

1 INTRODUCTION

Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC, stipulates in Article 4 (1) and Annex II paragraph 3.1 that for the approval of an active substance, where relevant, acceptable daily intake (ADI) and acceptable operator exposure level (AOEL) shall be established.

The ADI of a chemical is the estimate of the amount of a substance in food or drinking water, expressed on a body weight basis, that can be ingested daily over a lifetime without appreciable health risks to the consumer on the basis of all known facts at the time of the evaluation (WHO, 1997)[2]. Currently, although the general principles for setting an ADI are described in various publications (e.g. WHO, 1994; 2009)[3], there is no harmonised and internationally agreed guidance for setting the acceptable daily intake of pesticide active substances to assist the European Commission and Member States when making decisions about inclusion of an active substance in Annex I of Directive 91/414/EEC or under Regulation (EC) No. 1107/2009.

The ADI concept was introduced in 1957 by the Council of Europe and later on was taken over by the Joint FAO/WHO Meeting on Pesticide Residues (JMPR). To date the ADI is considered as being a valid tool in risk assessment (Galli et al., 2008). Traditionally, ADI values are based on reference points (RPs) derived from results of long-term animal toxicity studies. An uncertainty factor is usually applied to convert the relevant reference point, normally No Observed Adverse Effect Levels (NOAELs) or Lowest Observed Adverse Effect Levels (LOAELs), in the absence of relevant NOAELs, into a safe daily intake value for humans. In recent publications, the appropriateness of this approach has been questioned since the establishment of NOAELs/LOAELs is dependent on study design, dose selection, group size, criteria or definition of adversity and the precision with which the test is performed (Renwick et al., 2003). The Benchmark Dose (BMD) approach arose as an alternative way of defining reference points for risk assessment. It was concluded in a recent opinion of EFSA's Scientific Committee, that the BMD approach is a scientifically more advanced method as compared to the NOAEL approach for deriving a reference point, since it also includes consideration of the dose-response curve and quantifies uncertainty and variability of dose-response data (EFSA, 2009b).

Historically, the routinely applied default uncertainty factor (UF) of 100 was introduced in 1954 by Lehman and Fitzhugh. It is designed to reflect application of two separate 10-fold factors for inter-species differences (animal-human) and human variability respectively (WHO, 1987). The traditional inter-species factor of 10 was later divided into values of 4 for differences in kinetics and 2.5 for toxicodynamic differences (Renwick et al., 1993). The equivalent subdivision of the UF for intra-species variation was to 3.16-fold for both kinetics and dynamics Subdivision of the UFs permits the use of specific data on a chemical to derive chemical-specific adjustment factors (CSAF) (WHO, 1994, 2001). Chemical-specific data for one particular aspect of uncertainty should be used to replace the relevant part of the overall default uncertainty factor (WHO, 2001: 2005). In recent publications human variability in kinetics has been analysed for the main metabolic pathways in healthy adults and subpopulations of the elderly, neonates and children. Based on this analysis pathway-related UFs could be established that allow the incorporation of metabolism data into the derivation of health-based guidance values (Dorne et al., 2004). In a recent review four main scenarios were identified

[2] References cited in the tender specification document are not presented, but additional ones are described.

[3] WHO, 2009. International Programme on Chemical Safety: Principles and methods for the risk assessment of chemicals in food. Environmental Health Criteria, 240,

where the current default uncertainty factor for toxicokinetics (3.16) do not cover the human variability (Dorne, 2010). The appropriateness of the current approaches to deal with the uncertainty in the establishment of health-based guidance values should therefore be evaluated and, if relevant, other science-based approaches for the application and refinement of uncertainty factors in setting the ADI should be elaborated.

According to Directive 97/57/EC (establishing Annex VI, the uniform principles to Directive 91/414/EEC), the Acceptable Operator Exposure Level (AOEL) is the maximum amount of active substance, expressed on a body weight basis, to which the operator may be exposed without any adverse health effects. Thus the AOEL is a health-based exposure limit to be used for comparison with estimated or measured exposure levels of operators, workers and bystanders for assessing the risk of these groups arising from the application of a plant protection product. AOELs should normally be based on reference points (i.e. NOAELs, alternatively in some cases also LOAELs) that are obtained in short-term animal toxicity studies. Based on assumptions similar to those described for setting the ADI, a default uncertainty factor of 100 is also applied to derive an AOEL (EC, 2006). However, very similar considerations and criticisms, as described above for setting ADIs, are valid for the appropriateness of the way of setting the AOEL values (both for deriving reference points and uncertainty factors).

Moreover, EFSA's Panel on Plant Protection Products and their Residues (PPR) recommended, in a recently adopted opinion on Preparation of a Guidance Document on Pesticide Exposure Assessment for Workers, Operators, Bystanders and Residents, that guidance should be developed on the derivation of a new reference value, a so-called "acute AOEL" or "AAOEL", that is required for an acute risk assessment for operators, workers and bystanders for substances that are acutely toxic (EFSA, 2010). Therefore the existing Draft Guidance for the Setting and Application of Acceptable Operator Exposure Levels (AOELs) (EC, 2006) should be updated to provide guidance on how to derive and apply this new health-based guidance value (AAOEL).

In addition to the above, the Pesticide Steering Committee of EFSA recently identified the necessity of giving high priority to the development of a new guidance on setting the acceptable daily intake of pesticide active substances and the revision of the existing draft guidance for setting and application of acceptable operator levels.

To address these aspects, EFSA let a tender for which CRD was the successful bidder.

This report and its associated Annexes present the analyses performed as part of the project, together with conclusions and recommendations. The report is presented in order of the activities and tasks as described in the tender agreement, followed by the conclusions and recommendations. The main body of the report contains summary information and analyses. Further details are presented either as Annexes to the report or as separate files. The supporting Excel spreadsheets containing data used for the analyses presented in the main text are presented as separate searchable files. This permits 3rd parties to review the basic information and perform their own analyses.

Participants in the project

The primary CRD contributors were:

- Dr S Brescia
- Dr H McGarry
- Dr I Dewhurst

Additional contributions and discussions were received from other CRD specialists with relevant expertise. Dr A Hart (FERA, UK) provided advice on the uncertainty aspects.

A small expert panel was convened to provide external peer review of the work. The Panel met twice, on 29th August 2012 and 31st January 2013. The four independent members of the Panel were:

- Dr Marloes Buschers (NL)
- Professor David Coggon (UK)
- Professor Andrew Renwick (UK)
- Dr Roland Solecki (DE)

Following technical difficulties with the Benchmark Dose case-studies, EFSA granted a three-month extension to the project deadline. Professor Wout Slob (RIVM, NL) provided guidance and advice on the BMD analyses and performed the analyses presented in the final report.

2. EFSA TASK 1:

Systematically and comprehensively evaluate EFSA's conclusions on ADI and AOEL setting with an emphasis on the rationales used and generic expert discussions for the selection of critical study, derivation of the relevant reference points (i.e. NOAELs and LOAELs), and the derivation of uncertainty factors; also compare ADI values and derivations with those of JMPR.

2.1 Methodology

Copies of the EFSA conclusions on pesticide active substances were downloaded from the EFSA website[4]. The 224 EFSA conclusions on pesticide active substances published up to 1st December 2012 were evaluated to determine the basis for the derivation of the ADI and AOEL. These conclusions covered new active substances and those reviewed under the so-called List 2 and 3 procedures. The data for the following fields were compiled in Excel spreadsheets:

- date of the conclusion;
- value for the ADI or AOEL;
- NOAEL used to derive the ADI / AOEL and the associated LOAEL;
- value of the safety factor (SF) applied;
- reasons if the SF was not 100;
- end point(s) used to derive the ADI or AOEL;
- critical target organs or tissues;
- species used for the critical studies;
- critical study type / duration;
- whether the original proposal in the DAR was supported;
- the ratio of the LOAEL to the NOAEL;
- additional comments.

The spreadsheets were used to perform analyses on these aspects:

- ranges of ADIs / AOELs, NOAELs and safety factors;
- common study types, species and durations used to derive the reference values;
- ratio of LOAELs to NOAELs;
- distribution of NOAELs associated with particular end-points;
- whether the original proposals by the rapporteur member states were supported.

In a number of instances it was necessary to go back to the Draft Assessment Report (DAR) to gather the required information; this applied to nearly all instances for the LOAEL value and original RMS proposals. The DARs are available on the EU website 'CIRCABC'[5] or on request from EFSA[6].

[4] http://www.efsa.europa.eu/en/publications.htm?scdtype=conclusion
[5] https://circabc.europa.eu/faces/jsp/extension/wai/navigation/container.jsp?FormPrincipal:_idcl=FormPrincipal:_id3&FormPrincipal_SUBMIT=1&id=8bd5dd33-9ab6-4b8c-a925-e1970255153a&javax.faces.ViewState=rO0ABXVyABNbTGphdmEubGFuZy5PYmplY3Q7kM5YnxBzKWwCAAB4cAAAAAN0AAE1cHQAKy9qc3AvZXh0ZW5zaW9uL3dhaS9uYXZpZ2F0aW9uL2NvbnRhaW5lci5qc3A=
[6] http://dar.efsa.europa.eu/dar-web/provision

In addition to the EFSA conclusions, analyses were performed on compounds for which both EFSA and the JMPR had set references values (n=57). The JMPR summary information is available on its website[7] together with further details in the associated toxicological monographs[8]. The aim of this analysis was to determine what degree of consistency there is between the two organisations and, if there are differences, what are the reasons. In addition, the derivations of medium term AELs (Acceptable Exposure Levels; equivalent to AOELs) for those biocide active substances authorised in the EU that are also pesticides were analysed.

Furthermore, the main discussion points identified during the written commenting phase and the 'EPCO' and 'PRAPeR' peer review meetings were evaluated to determine if there were any topics that routinely presented difficulties and should be prioritised when developing future guidance. The basic information in the form of commenting and evaluation tables and meeting reports is available from 'CIRCABC'[9].

The supporting data for the analyses summarised below are presented in a number of spreadsheets attached to the report as Excel files. The contents of the supporting spreadsheets are explained in Annex 1.

The overall spreadsheets also contain some information from the so-called List 4 review compounds. Initially, these compounds were included in the summary analyses but they have now been removed. The list 4 compounds typically have limited databases, because they are naturally occurring or their main use is not as a pesticide. Therefore, the derivation of the ADI or AOEL for these substances often requires approaches that are not routinely used for active substances in other lists/groupings and the patterns of results are inconsistent with those of synthetic, chemical active substances.

2.2. Rationale for conclusions

The rationale for the choice of individual NOAELs and ADIs or AOELs was not always clearly described in the documentation. This was particularly so when the standard or default approaches (e.g. as described in WHO, 2009), with statistically significant findings treated as adverse and a SF of 100, were used. More details of the underlying rationales were normally, but not always, provided when non-default SFs were used, findings were set aside as non-relevant to humans or combined NOAELs were derived from two or more studies. Information on the rationales is provided in the spreadsheets and key aspects are highlighted in the summary texts.

2.3. Conclusions on Acceptable Daily Intakes

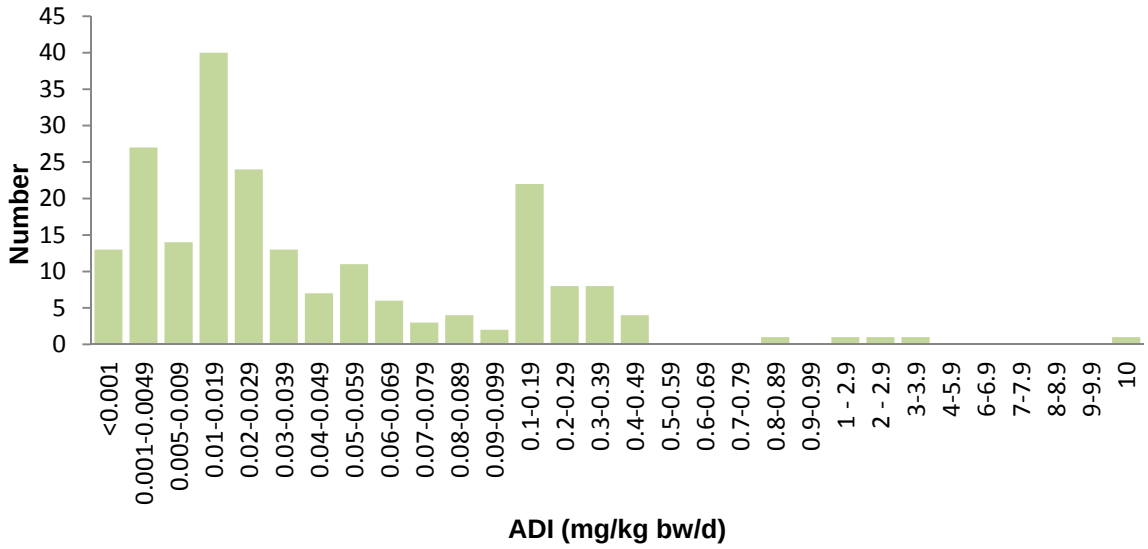
2.3.1. ADI analyses for EFSA compounds (n=224; values were set for 214)

Within the analyses it should be noted that the sum of individual values can equal more than the total number of compounds / ADIs set, as for some evaluations more than one study was used as the basis for the ADI.

[7] http://www.who.int/entity/foodsafety/chem/jmpr/publications/pesticide_inventory_report_2010.pdf
[8] http://www.who.int/foodsafety/chem/jmpr/publications/monographs/en/index.html
[9] https://circabc.europa.eu/faces/jsp/extension/wai/navigation/container.jsp?FormPrincipal:_idcl=FormPrincipal:left-menu-link-lib&FormPrincipal_SUBMIT=1&javax.faces.ViewState=rO0ABXVyABNbTGphdmEubGFuZy5PYmplY3Q7kM5YnxBzKWwCAAB4cAAAAAN0AAE4cHQAKy9qc3AvZXh0ZW5zaW9uL3dhaS9uYXZpZ2F0aW9uL2NvbnRhaW5lci5qc3A=

Range of ADI values (mg/kg bw/day)

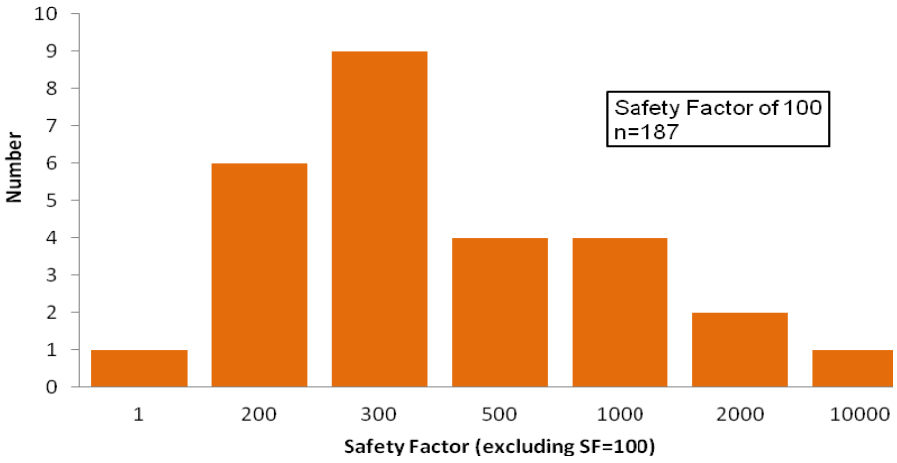
0.00015 – 10 Mean 0.14 Median 0.023



Safety factor not 100

N= 26

Range 1 – 10,000

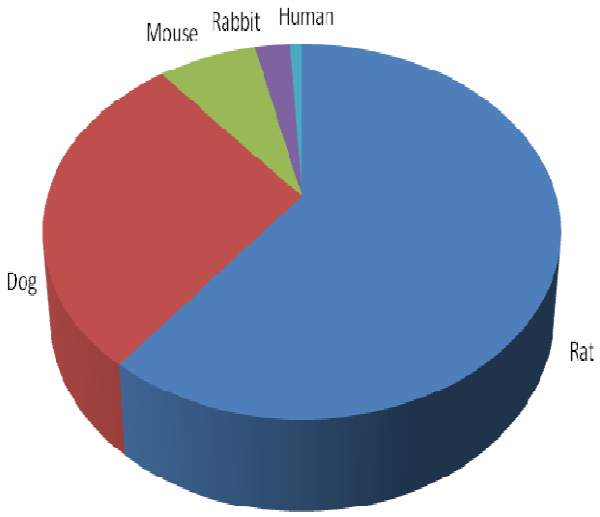


Main Reasons:

- Provide margin of safety to severe effects 4
- Use of LOAEL 14
- Severity / tumours 6

The factor of 10,000 was set for 1-methylcyclopropene, a gas, and included additional factors for an absence of chronic studies and a correction for 10% absorption via the respiratory tract.
The factor of 1 was for copper compounds and was based on data from human dietary exposures.

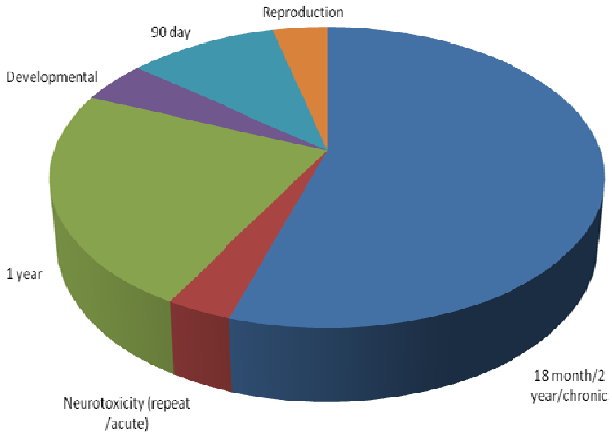
Frequency of species used to set ADI



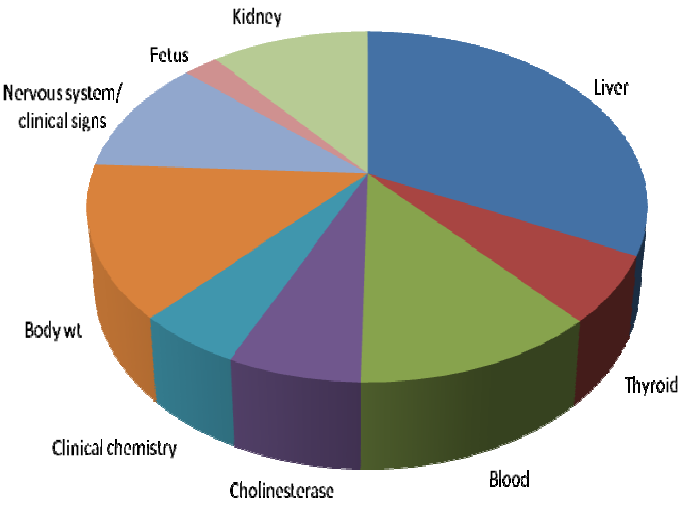
Rat	62%
Dog	31%
Mouse	7%
Rabbit	2%
Human	21%

Frequency of main study types used to set ADI

18 month / 2 year / chronic	64%
1 year	26%
90 day	10%
Developmental	4%
Reproduction	6%
Neurotoxicity (repeat /acute)	4%



Main targets / end-points



Liver	43%
Thyroid	8%
Blood	16%
Cholinesterase	9%
Clinical chemistry	7%
Body wt	11%
Nervous system / clinical signs	13%
Fetus	3%
Kidney	14%

No value set

N= 10

Main Reasons

- Inadequate data / missing information 5
- No exposure / within background 3

2.3.2. Comparison with values derived by JMPR

Although the overlap of substances evaluated by both EFSA & JMPR was not extensive, there were 57 compounds that had been reviewed by both organisations. A comparison of the derivation of the ADIs was performed and is summarised in Table 1. The ADI values concluded by EFSA and JMPR were generally in broad agreement. Discrepancies were not accounted for by a single explanation; rather, a variety of scenarios was apparent. The most common of these was the choice of study from which the NOAEL was chosen. Additionally, the JMPR occasionally used NOAELs from several studies to derive the ADI, albeit NOAELs that were very similar. Occasionally, the two conclusions applied different safety factors, although, like EFSA, the default safety factor of 100 was most commonly employed by JMPR. In rare cases, either EFSA or JMPR had decided not to set an ADI, either because of inadequate data or because it was not deemed necessary (no residue in food), whereas the other organisation had set a value. Some of the JMPR conclusions dated from the 1990s or earlier and had set ADI values that were higher than the more recent EFSA conclusions. In only two instances was there clear evidence of critical studies being available to one group but not the other.

For 34 of the 57 compounds (60%) the values were the same apart from rounding differences. Of the 22 values that differed, the JMPR value was higher in 13 cases (60%).

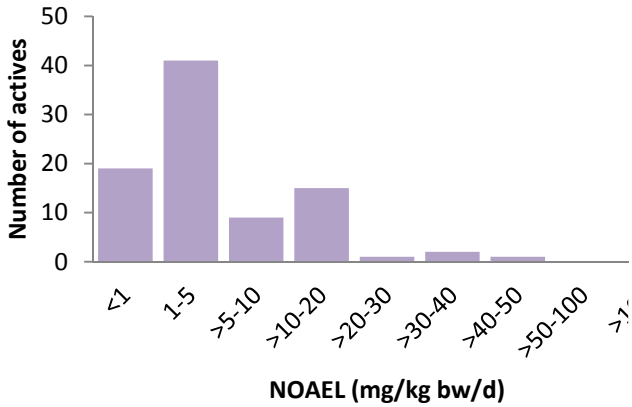
*Table 1 – Reasons for differences between JMPR & EFSA ADIs**

Reason	**2. Number**	**Active substances**
Same value, different approach	1	Abamectin
Use of overall / combined NOAEL (used by both organisations, not always consistently)	4	Bitertanol, Dimethomorph, Myclobutanil, Triadimenol
Different study / approach	9	Cadusafos, Carbendazim, Diazinon, Diflubenzuron, Dimethoate, Fenbuconazole, Methyl bromide, Propamocarb, Thiodicarb,
Use of human data by JMPR	3	Ethephon, Methomyl, Pirimiphos-methyl,
Used study not available to other group	2	Carbofuran, Carbosulfan
Filling of large dose spacing (both groups have filled dose spacing for other compounds)	2	Clethodim (75x), Thiomethoxam (25x)
Different additional safety factor	1	Dicloran
Conversion from dietary level to dose	1	Chlormequat
Same value	26	Buprofezin, Captan, Clofentezine, Cycloxydim, Cyprodinil, Cyromazine, Dodine, Ethoprophos, Etofenprox, Fenamiphos, Fenpropimorph, Fenpyroximate, Fipronil, Fluopicolide, Flutolanil, Folpet, Glufosinate, Hexythiazox, Kresoxim methyl, Phosmet, Prochloraz, Prothioconazole, Pyriproxifen, Tebuconazole, Tebufenozide, Teflubenzuron.
Same value apart from rounding (JMPR normally rounds – EFSA uses 2 significant figures)	8	Bifenthrin, Carbaryl, Diphenylamine, Haloxyfop, Imazalil, Pyrimethanil, Spirodiclofen, Sulfuryl fluoride,
3. Total	57	

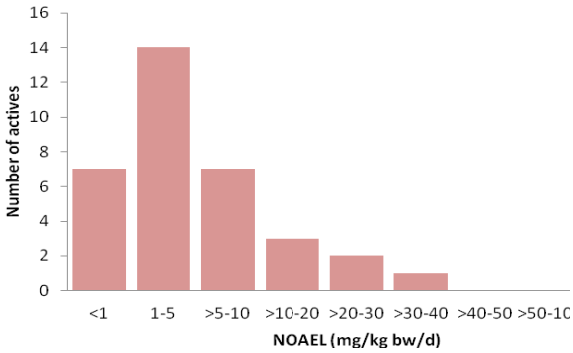
*excludes those where no value was set.

Analysis of NOAELs for particular end-points used for setting ADIs

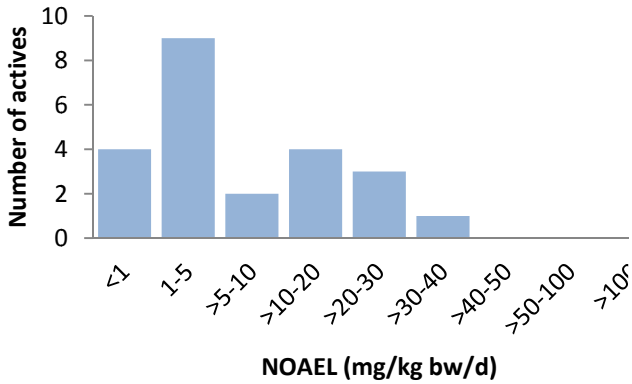
ADI liver



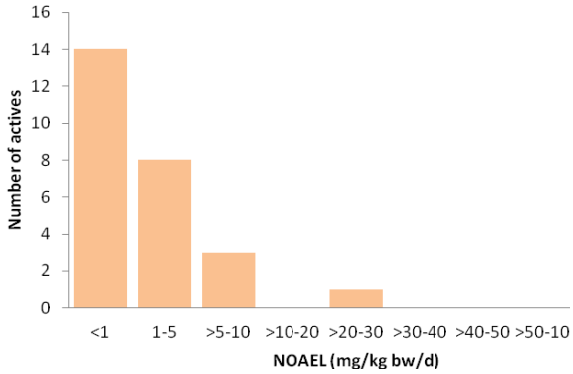
ADI blood



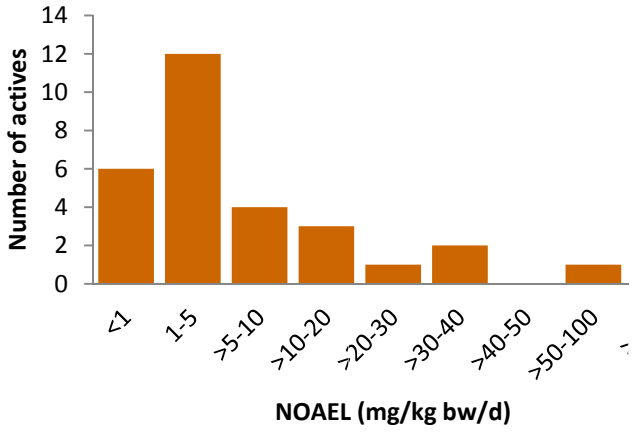
ADI bd wt



ADI signs



ADI kidney



These plots for the NOAELs for liver effects, reduced body weight, clinical signs, kidney lesions and haematological effects indicate that, other than for clinical signs, there is a broad range of potencies

and none of the end-points can be considered as occurring only at either relatively high or relatively low dose levels.

Analysis of LOAEL to NOAEL ratio

For the ADIs set by EFSA, the ratio of LOAEL to NOAEL for the critical study(ies) varied from 1.5 to 225 (for diazinon, 0.02 to 4.5 mg/kg bw/day). For 23 compounds the ratio was 2 or less, but 19 compounds had ratios in excess of 10. The mean value was 8.2, which is probably distorted by a small number of large values. The median value is 5. This analysis shows that for many pesticide active substances the uncertainties around the ADI could have been reduced if tighter dose spacing at the low end of the dose-response curve had been incorporated in the study design.

Analysis of original DAR proposals

For 128 compounds (57%), the original proposal by the RMS was in agreement with the final ADI value. Reasons for changing the proposal included different safety factors, new data submitted during the procedure and a different NOAEL chosen.

Summary

The analysis of 224 pesticide active substances with EFSA conclusions and ADI considerations produced results broadly in line with expectations.

- For the majority of compounds (57%) the original proposal made in the DAR was supported through the peer review.
- The majority of ADIs were based on chronic studies (18 months or longer duration) as would be expected for a lifetime-exposure-based guidance value.
- Studies in rats were the most frequently used, followed by dogs. Mouse data were used in 7% of cases but often with other studies.
- The default safety factor of 100 was used in the majority of cases (*ca* 75%). Where a factor other than 100 was used this was most frequently to correct for the use of a LOAEL (n=14). In only one case (carbofuran) was any effort (BMD) made to support the magnitude of the extra factor.
- General end-points (e.g. liver enlargement / hypertrophy and lower body weight) were utilised in setting the majority of ADIs.
- The EFSA and JMPR conclusions were generally in at least broad agreement. The most common reason for discrepancies was the choice of study on which to choose the NOAEL used to derive the ADI.
- The range of values in this sample is consistent with that of the broader EU database (0.0008 – 10 mg/kg bw/day[10]) and there is no reason to believe the omission of pre-EFSA evaluations has affected the general findings.
- For a significant number of compounds (>40%) the dose spacing between the LOAEL and NOAEL in the critical study was greater than 5. There is therefore considerable potential to reduce the uncertainty surrounding the ADI, or to refine it, by more consideration of dose selection at the lower end of the dose response curve.

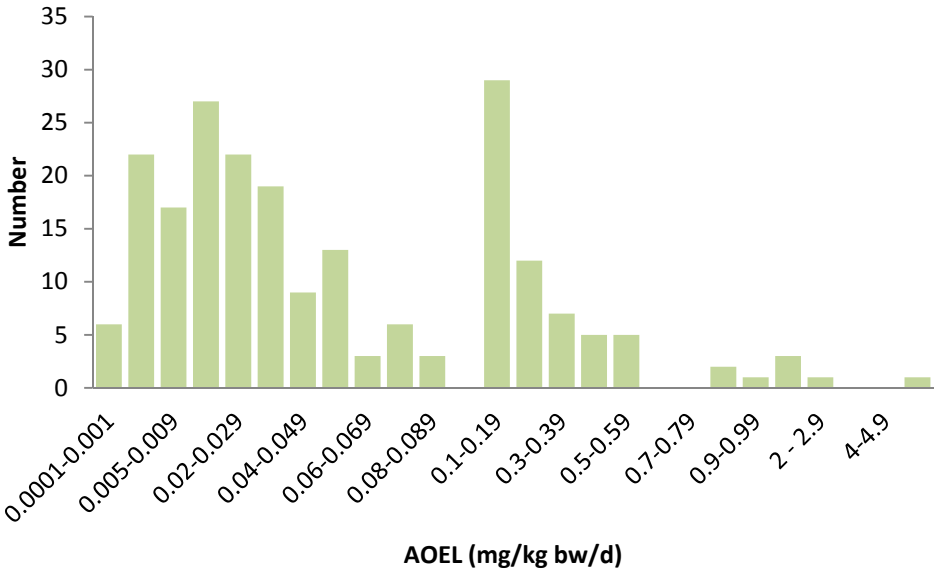
[10] http://ec.europa.eu/sanco_pesticides/public/index.cfm?event=activesubstance.selection

2.4. Conclusions on Acceptable Operator Exposure Levels

2.4.1. AOEL (n=224; values set for 215)

Range of values (mg/kg bw/day)

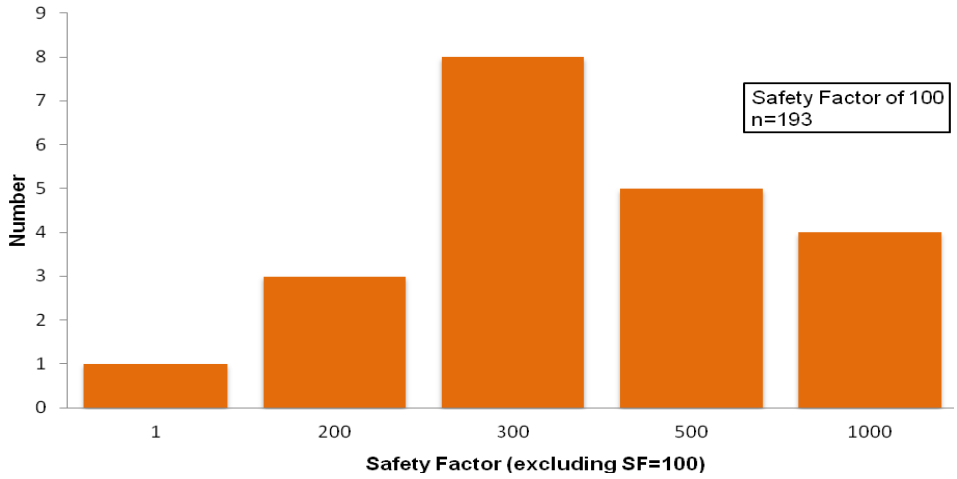
0.0002 – 5.0 Mean 0.14 Median 0.03



Safety factor not 100

N= 20

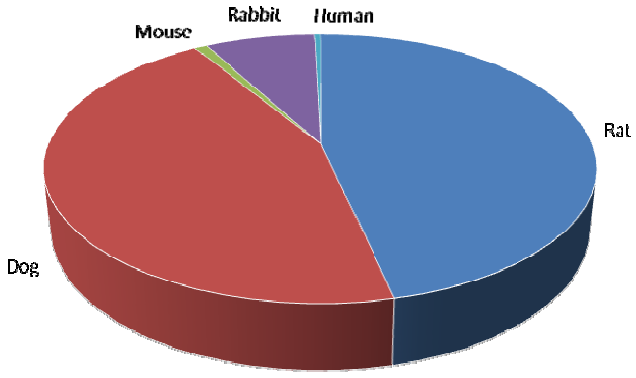
Range 1 – 1,000



Main Reasons:

- Human data 1
- Lack of / missing data 5
- Provide margin of safety to severe effects 3
- Use of LOAEL 6
- Severity / tumours 5

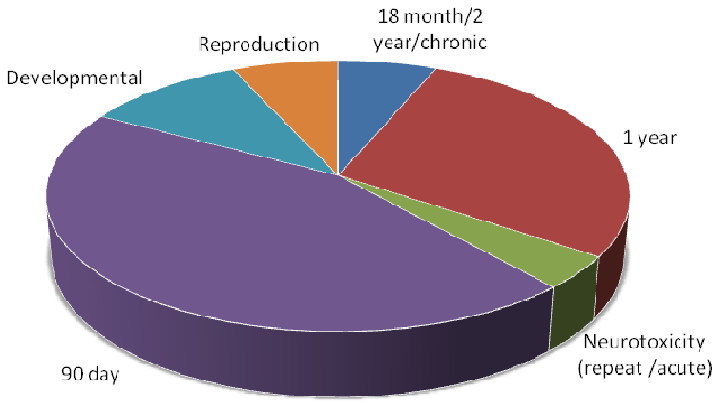
Frequency of species used to set AOEL



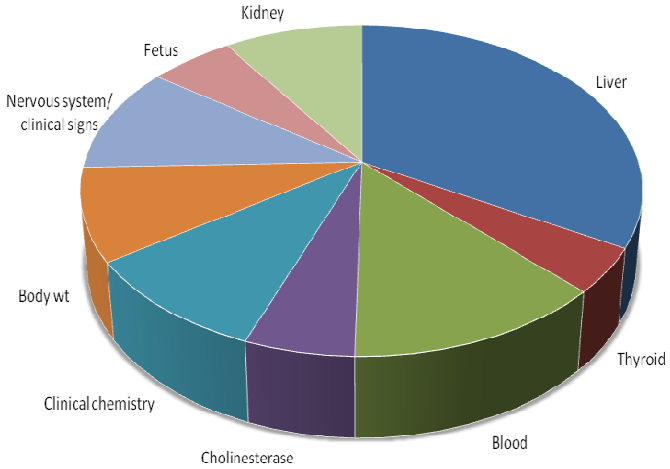
Rat	45%
Dog	48%
Mouse	1%
Rabbit	6%
Human	0.5%

Frequency of main study types used to set AOEL

18 month / 2 year / chronic	7%
1 year	34%
90 day	49%
Developmental	10%
Reproduction	7%
Neurotoxicity (repeat /acute)	5%



Main targets / end-points



Liver	44%
Thyroid	6%
Blood	4%
Cholinesterase	7%
Clinical chemistry	11%
Body weight	8%
Nervous system/clinical signs	13%
Fetus	7%
Kidney	12%

3.1.1. No value set

N= 9

Main Reason

- Inadequate data / missing information 6

Oral absorption correction

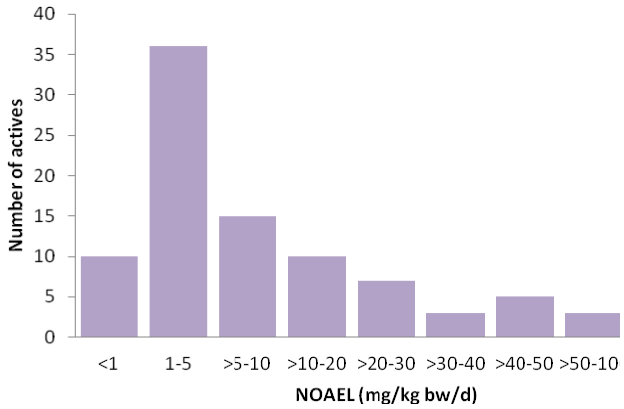
N= 72 (33%)

Range 8 – 80%

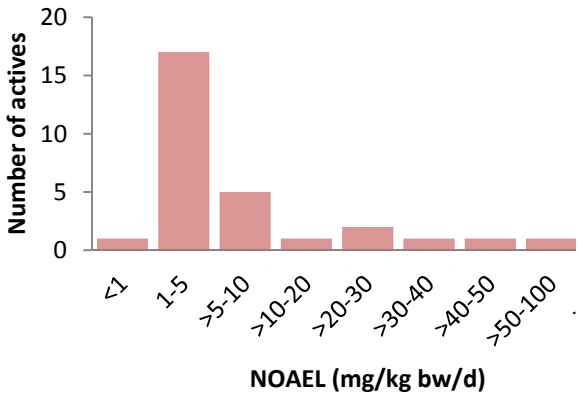
There was inconsistency about correcting values near 80%. Flutolanil was corrected by 80%; but terbuthylazine was not corrected for 79%.

Analysis of NOAELs for particular end-points used for setting AOELs

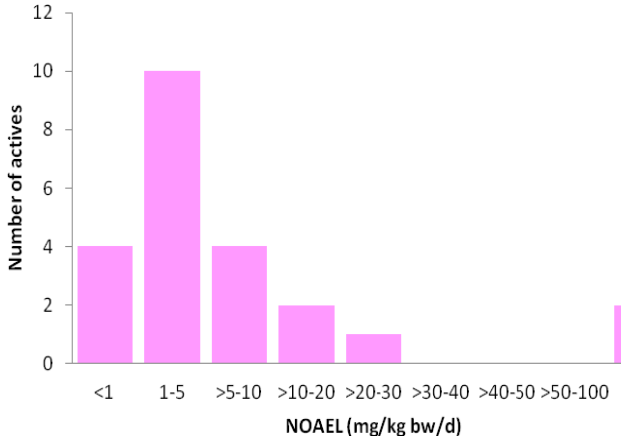
AOEL liver



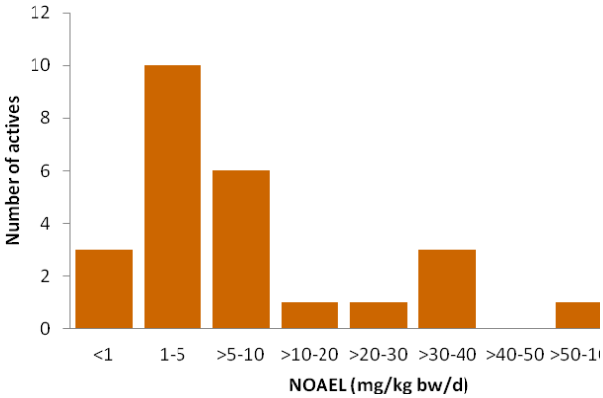
AOEL blood



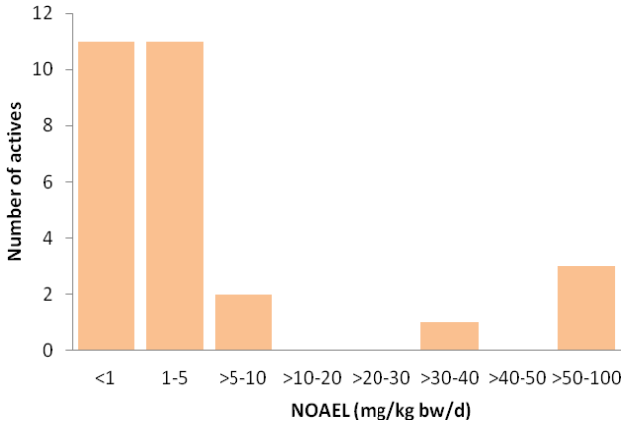
AOEL clin chem.



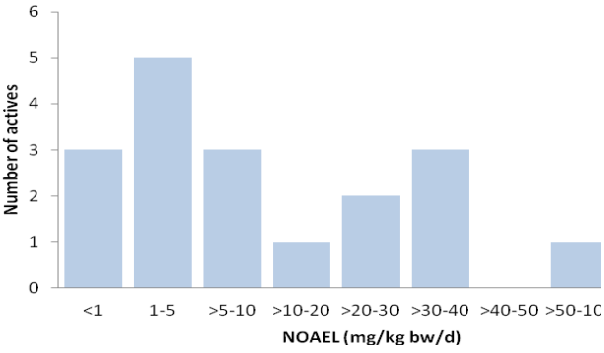
AOEL kidney



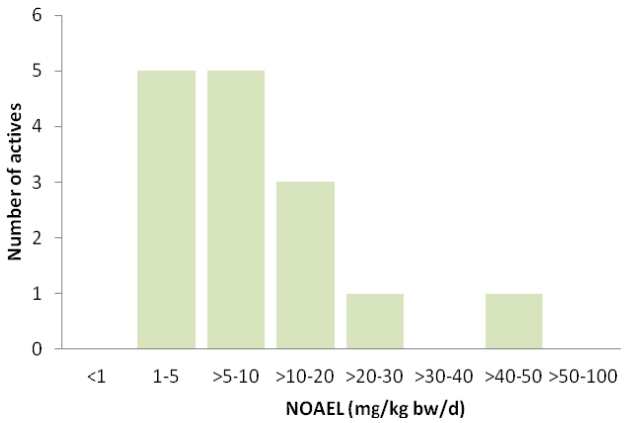
AOEL nerve



AOEL bd wt



AOEL foetus and offspring



These plots of the NOAELs for liver effects, fetal and offspring effects, reduced body weight, clinical chemistry changes, nervous system /clinical signs, kidney lesions and haematological effects indicate that there is a broad range of potencies and none of the end points can be considered as being highly sensitive.

Analysis of LOAEL to NOAEL ratio

For AOELs set by EFSA, the ratio of LOAEL to NOAEL for the critical study(ies) varied from 1.4 to 225 (for diazinon, 0.02 to 4.5 mg/kg bw/day). For 40 compounds the ratio was 2 or less but 15 compounds had ratios in excess of 10. The mean value was 6.4, which is probably distorted by a small number of large values. The median value is 4. This analysis shows that, for many pesticide active substances, the uncertainties around the AOEL could have been reduced if tighter dose spacing at the low end of the dose-response curve had been incorporated in the study design.

Analysis of original DAR proposals

For 108 compounds (48%), the original proposal by the RMS was in agreement with the final AOEL value. Reasons for changing the proposal included different safety factors, different oral absorption correction, new data submitted during the procedure, alternative study duration to match the use pattern of the pesticide and a different NOAEL chosen.

Use of 1 year dog data

One surprising aspect of the assessment was that 1 year dog studies were used for setting a significant number of ADIs and AOELs. This was not expected give recent considerations of the value of the 1 year dog study and its removal from the EU data requirements. To investigate this further the 1 year and 90 day dog studies for 45 active substances (48 evaluations), where the AOEL was based on the 1 year study, were compared to see if the 1 year studies were giving NOAELs clearly below those from the 90 day studies. The results of this analysis were:

- For 19 of the compounds no acceptable 90 day dog study was available;
- For 12 of the compounds the 90 day NOAEL was at or below the 1 year LOAEL, indicating the apparent differences in sensitivity might be due to dose spacing;
- For 6 of the compounds the 90 day NOAEL was below the 1 year NOAEL;

- For cyromazine, the critical effects (haematology) were first seen at the 8 week sampling time and it is unclear why they were not seen in the 90 day study;
- For metam, the changes considered adverse in the 1 year study were not considered adverse in the 90 day study (e.g. ALT 279% in 1 yr = LOAEL and 371% in 90 day = NOAEL);
- For the Haloxyfops, the RMS had proposed a NOAEL for the 90 day study at 2 mg/kg bw/d and for the 1 year study 5 mg/kg bw/d; the 1 year NOAEL was lowered to 0.5 at peer review.
- For 4 compounds (Acequinocyl, lufenuron, pyriproxifen and tebuconazole) the indications were that the effects seen in at the LOAEL for the 1 year study were not evident in the 90 day study at the same or slightly higher dose level. However there were some methodological differences between the studies and more investigation of the original study reports is required to reach a robust conclusion on the value of the 1 year study for these compounds.

Based on this analysis the 1 year study appears to be of value to the risk assessment for less than 10% of the compounds where it is used to set the AOEL. This is consistent with the previous analyses of the value of a 1 year dog study and supports the conclusion that a 1 year dog study should not be a routine requirement.

2.4.2. Comparison of the pesticide AOEL with the biocide medium-term AEL

Among the 60 active substances included in Annex I of the Biocidal Product Directive (BPD), 17 have also been assessed for pesticidal use (with EFSA conclusions available). For these 17 substances, the systemic AOEL set within the pesticide regime has been compared with the equivalent systemic medium-term AEL (Acceptable Exposure Level) established within the biocide scheme (see spreadsheet 7 in Annex 1).

The analysis has shown that for 12/17 (70%) substances, the AOEL has the same value as the medium-term AEL and has been established from the same study, same starting point, same assessment factors and same oral absorption value. For the remaining 5/17 (30%) substances (copper compounds, difenacoum, etofenprox, imidacloprid, tolyfluanid), the values differ. For 3 of these 5 substances (copper compounds, etofenprox, imidacloprid), the medium-term AEL is slightly higher than the AOEL; for 1 substance (tolyfluanid), the medium-term AEL is slightly lower than the AOEL; and for 1 substance (difenacoum), the medium-term AEL is more than one order of magnitude lower than the AOEL.

The main reason for the different values is a different starting point from a different study (copper compounds, difenacoum, imidacloprid, tolyfluanid). This could be because of differences in the datasets or to a different interpretation of the same datasets. In one case (etofenprox), a different oral absorption value was used. For difenacoum, for which the medium-term AEL is more than one order of magnitude lower than the AOEL, a lower starting point from a different study (not available to the pesticide regime) was used for biocides, combined with a higher overall assessment factor (owing to the fact that the starting point was a LOAEL) and a lower oral absorption value.

Approximately 47% (8/17) substances have been assessed under the biocide scheme a few years later than under the pesticide regime; 4/17 (24%) substances have been finalised at a similar time under

both regimes; and 5/17 (29%) have been evaluated under the biocide scheme one or two years earlier than under the pesticide regime.

Overall, there is good agreement between the two legislative regimes, although for a significant minority of substances (30%), differences are apparent. The most prevalent reason for the observed differences appears to be a slightly different dataset, with several studies available to one regime but absent from the other.

2.4.3. Summary of AOEL analyses

The analysis of 224 pesticide active substances with EFSA conclusions and AOEL considerations produced results broadly in line with the existing guidance document.

- For the slightly less than half (48%) of compounds the original proposal made in the DAR was supported through the peer review. This is a lower proportion than for the ADI and relates to additional considerations associated with the AOEL (oral absorption and appropriate study duration).
- The majority of AOELs were based on 90 day studies (n=106) but 1 year studies were used in a relatively large number of cases (n=73).
- Although the 1 year dog study was used for 45 active substances for only 4 of these was there evidence that the 1 year study was more sensitive than the 90 day study.
- The guidance document indicates that studies of durations up to and including 90 days would normally be used for AOELs. Fifteen AOELs were set on studies with a duration of 18 months or longer. However, for the majority of these the longer-duration studies were used only in conjunction with studies of shorter duration. For the three compounds where chronic studies were used alone (2-phenylphenol, bupyrimate and zeta cypermethrin), the reasons for the use of the longer duration study are unclear.
- Studies in rats and dogs were used equally.
- The default safety factor of 100 was used in the majority of cases (*ca* 85%). Where a factor other than 100 was used, this was most frequently to correct for the absence of data or the severity of effects.
- A correction for oral absorption was made in approximately one third of AOELs. The use of the 80% cut-off for oral absorption correction introduced some significant discrepancies (e.g. for terbuthylazine a correction was made for 79% absorption, but not for several compounds with oral absorption of 80% or *ca* 80%).
- There was no obvious pattern to when data from reproduction studies were used in deriving the AOEL (rarely) or when 1-year or longer studies were taken into account (one third of cases). Additional information is required regarding the working patterns of pesticide operators, including contractors so that the appropriate duration of study can be identified and agreed for particular activities.
- For 4 compounds route specific or chronic AOELs were set.
- For a significant number of compounds (>25%) the dose spacing between the LOAEL and NOAEL in the critical study was greater than 5. There is therefore considerable potential to reduce the uncertainty surrounding the AOEL, or to refine it, by more consideration of dose selection at the lower end of the dose response curve.

- There was consistency between the biocide and pesticide evaluations in the majority of cases.
- The values in this sample are consistent with the broader EU database (0.000017 – 128 mg/kg bw/day[11]); the latter value is for potassium bicarbonate, the next highest is 14 for imazamox.

2.5. Analysis of pesticide active substances whose reference doses are based exclusively on effects on body weight and/or liver weight

Among the 224 pesticide active substances for which EFSA conclusions have been considered further in this project (see supporting data in spreadsheet 1 in Annex 1), 32 (14%) were identified as having at least one reference value set exclusively on the basis of effects (at the LOAEL) on body weight and/or liver weight (see spreadsheet 8 in Annex 1).

Among these 32 substances, the reference value concerned was the ADI for 17 substances; the AOEL for 14 substances; and the ARfD for 13 substances. For some substances (e.g. folpet, picloram, dodine, pyridaben, prosulfocarb, etc.), two or even all three reference values were based on such effects.

2.5.1. ADI

For the 17 substances for which the ADI was based on such findings, 14 caused effects on body weight, 2 on liver weight (diethofencarb, hymexazol) and 1 on both body and liver weight (kresoxim-methyl). Of these 17 substances, there were 11 where the effects concerned were observed in the 2-year chronic study in the rat; 4 where the effects concerned were noted in the 1-year/2-year study in the dog; and 2 where the effects concerned were reported in the multi-generation study in the rat.

Reductions in body weight ranged from 4 to 24% of the control values. Reductions in body weight gain were higher, ranging from 15 to 52% of the control values. For the majority of the substances, such reductions were statistically significantly different from controls. The substance with the lowest decrease in body weight (4%) was chlorsulfuron. It is unclear from the information available in the DAR whether this decrement, which was seen in males only from the rat chronic study, was statistically significant. Only 2 (chlorsulfuron and amidosulfuron) out of 14 substances had reductions in body weight < 10%. Reductions in body weights were associated with reductions in food consumption in approximately 2/3 of the substances considered.

For 5 out of the 17 (29%) substances considered, no significant toxic effects other than reductions in body weight were observed up to the highest doses tested. In the remaining substances, additional toxic effects were noted at higher dose levels.

For the 3 substances with increased (absolute) liver weight, the magnitude of the increase (24%, statistically significant) was specified in the DAR only for one (diethofencarb).

2.5.2. AOEL

For the 14 substances for which the AOEL was based on such findings, 11 caused effects on body weight and 3 on dog liver weight (etridiazole, hymexazol, isoxaben). Of these 14 substances, there were 6 where the effects concerned were observed in the 90-day/1-year study in the dog; 5 where the

[11] http://ec.europa.eu/sanco_pesticides/public/index.cfm?event=activesubstance.selection

effects concerned were noted in the 90-day/multi-generation study in the rat; and 3 where the effects concerned were reported in the developmental toxicity study in the rabbit.

Reductions in body weight ranged from 6 to 18% of the control values. Reductions in body weight gain were higher, ranging from 15 to 85% of the control values. For the majority of the substances, such reductions were statistically significantly different from controls. The substance with the lowest decrease in body weight (6%) was diflufenican. This decrease was statistically significant and was observed in males only from the 90-day study in the rat. No other substance (out of the 11 identified) had reductions in body weight < 10%. Reductions in body weights were associated with reductions in food consumption in all except 1 of the substances considered.

Only for 2 out of the 14 (14%) substances considered, no other significant toxic effects than reductions in body weight were observed up to the highest doses tested. These were studies conducted in the dog. In the remaining substances, additional toxic effects were noted at higher dose levels.

For the 3 substances with increased (absolute) liver weight, the magnitude of the increase ranged from 14 to 26% of the control values. Statistical significance was not specified for 2 substances and was not attained for the third substance (isoxaben).

2.5.3. ARfD

For the 13 substances for which the ARfD was based on such findings, 10 caused effects on maternal body weight in developmental toxicity studies in rabbits and/or rats; 2 caused initial effects on body weight in the adult dog in 90-day or 1-year studies (bitertanol, fluoxastrobin); and 1 caused effects on foetal body weight in developmental toxicity studies in the rat (oxadiazon). None of these 13 ARfD values were based on effects on liver weight.

Reductions in body weight ranged from 1.4 (metamitron) to 8% (pyridaben) of the control values. Reductions in body weight gain were higher, ranging from 17 (prosulfocarb) to 86% (quinmerac) of the control values. For 3 substances (picloram, triflusulfuron and triticonazole), there was severe body weight loss occurring in maternal animals during the first days of dosing. For the majority of the substances, such reductions/losses were statistically significantly different from controls. Reductions in body weights were associated with reductions in food consumption in all of the 13 substances considered.

Only for 2 out of the 13 (15%) substances considered, no other significant toxic effects than reductions in body weight (in maternal animals of developmental toxicity studies) were observed up to the highest doses tested. However, even for these two substances, this pattern of effects is not considered unusual because investigations of maternal toxicity in prenatal developmental toxicity studies are very limited. In the remaining substances, additional toxic effects were noted at higher dose levels.

It has been proposed that the dose descriptors on which the ARfD values are established should also be used to derive the AAOEL (Acute Acceptable Operator Exposure Level). We note that for at least 10 substances (6%) out of the 161 for which an ARfD has been set, the dose descriptor is based on reductions in body weight of pregnant rats or rabbits treated by gavage. It is most likely that these effects are at least in part the consequence of the method of administration (gastric intubation) rather than a purely specific toxic effect of the chemical. It is therefore questionable whether such effects would be relevant to an operator exposed acutely to the substance by the dermal and inhalation routes and hence to the derivation of an AAOEL. However, if they are used as the basis of a first tier determination of an AOEL the outcome will be protective.

2.5.4. Conclusions

The relevance of moderately reduced body weight or body weight gain, particularly when it is transient, to humans is debatable and whether it is truly averse would depend on a number of factors.

- With current reports of increasing obesity in developed countries it could be argued that reductions in body weight gain could be positive rather than adverse. Although it is not normally considered adverse, there is no reason why increased body weight in laboratory animals should not be considered adverse.
- Laboratory rodents are provided with food *ad libitum* and minimal opportunity to exercise so are in an unnatural environment and (particularly in chronic studies) are significantly heavier than wild animals. Those in treated groups having lower body weights typically have greater survival at 18 – 24 months combined with lower levels of spontaneous lesions; therefore the lower body weight is beneficial.
- If the effect is in a dietary-exposure study and is secondary to palatability it is unlikely that humans would be similarly affected as the exposure would normally be >100-fold lower and thus less 'repulsive'. Additionally, humans have a less sensitive olfactory system than most other mammals.
- Food consumption of individual rodents in shared cages is difficult to determine, making it difficult to relate the food consumption of individuals with low body weight. In dogs, normally fed a defined mass of food per day, it is easier to investigate a relationship between food consumption and body weight.
- If the mode of toxic action is related to cellular energy production (e.g. some pesticides disrupt mitochondrial function) it would be reasonable to have a default assumption that any body weight losses or deficitswere a direct result of the mode of action and consider them as adverse.
- Animals cannot indicate that they 'feel unwell'. One result of being unwell might be reduced food intake and body weight gain. If the reason for being unwell is systemic then the body weight effects could be a sensitive marker of adversity and relevant to humans. This would also apply to compounds with a pharmacological action to reduce appetite.

For liver weight, the magnitude of effects at the NOAEL is consistent with the results of the evaluation of discussion points at peer review where there is a broad acceptance of a 10% increase in liver weight with associated changes in histopathology or clinical chemistry being considered as adverse.

2.6. Endpoints not routinely investigated

Toxicity test guidelines require the investigation of only certain end-points. It is not possible to investigate absolutely every possible effect a chemical might cause and current guidelines concentrate on the major organ systems and tissues. The possibility that certain sensitive end-points might be missing from the analyses described previously was considered. It was identified that certain end-points were used only very rarely in deriving health-based guidance values, e.g. cardiovascular effects, developmental neurotoxicity and immunotoxic effects. One of the reasons for this could be that specific investigations of certain end-points were not part of the routine OECD test guidelines or

the tests were not part of the standard regulatory packages. The key question is whether the absence of these studies or end-points would compromise the overall risk assessment.

An extensive evaluation of developmental neurotoxicity (DNT) studies submitted to the USEPA reported that investigations specific to DNT were critical to the risk assessments for only 4 of 75 pesticides tested (Makris et al, 2009). The absence of a DNT study was not considered significant to the risk assessment of chemicals. Basic indicators of immunotoxicity such as leucocyte counts and thymus pathology are present in routine test guidelines and have identified immunotoxic agents such as tributyltin (WHO, 1990). Specific investigations such as IgM response or response to parasites have confirmed an immunotoxic mode of action but not necessarily altered the NOAEL used to set the reference doses.

Cardiovascular system (CVS) effects other than heart weight and cardiac pathology are not routinely measured in toxicity studies. The OECD test guidelines do not require any investigations of pulse, blood pressure or electrocardiography, although such investigations can be performed easily, particularly in dogs. Significant cardiotoxicity would be evident in terms of pathological changes to the heart, heart weight, behavioural changes or death (Bharadwaj, 2009); if these are not evident at high dose levels, it is unlikely that significant changes in blood pressure or QT intervals would be seen at doses relevant to the NOAEL that are one or more orders of magnitude lower. Persistent high blood pressure can produce secondary effects in other organs such as the kidney or retina that form part of the routine examinations. What is not determined by such investigations is transient effects that might be of relevance to those with a pre-existing condition or working with machinery. Although it might not be relevant to all compounds, the inclusion of basic CVS measures could be something for consideration in any revisions to OECD test guidelines for repeat dose tests.

2.7. Analysis of reports from EPCO and PRAPeR meetings and associated teleconferences (TC) and written comments

As part of the EU procedures for assessing pesticide active substances, there is a peer review process. Since 2002, EFSA has been responsible for running the peer review process via a combination of meetings (initially called EPCO, latterly PRAPeR), telephone conferences and written commenting phases. All Member States can participate in the written commenting phase, meetings and telephone conferences; the notifying company can participate during the written commenting phase.

To determine if any particular aspects were regularly causing difficulties during peer review, an analysis was performed on the records of the peer review discussions. The information for the analyses of the discussions was obtained from records held on the EU CIRCABC website[12]. The main discussion items identified during the peer review process are presented in the attached spreadsheet (See supporting data in spreadsheet 2 in Annex 1).The spreadsheet covers discussions of the mammalian toxicology of 222 active substances and is divided into columns for the most common aspects (use of historic control data; classification and labelling; liver effects; body weight; safety factor and margin to LOAELs; oral absorption value applicable to the AOEL). The level of detail in the reports varied considerably and it was not always possible to determine the conclusions or the reasoning associated with the discussions. The fact that a topic is not mentioned in the records does not mean there were no discussions, but it is assumed that major discussion items would be recorded.

Common discussion topics from PRAPeR and EPCO peer review meetings:

[12] https://circabc.europa.eu/faces/jsp/extension/wai/navigation/container.jsp

- **Impurity profile.** The majority of substances had discussions regarding the relevance of the batches used in the toxicology tests to the commercial material as sold, with the impurity profile defined by the agreed technical specification. Reasons for this included: the absence of data on some or all of the toxicology batches; particularly pure material used for key studies such as genotoxicity; changes in the technical specification during the evaluation process.
- **Use of historic control data (HCD).** Historic control data can be useful in determining the relevance of findings in some toxicology studies. The use and applicability of the available HCD was discussed for 32 active substances (14%). Aspects appearing frequently included the relevance of data from generic databases; HCD being made available late on in the process.
- **Classification and labelling.** Significant time was spent discussing hazard identification / classification and labelling (71 active substances, 30%) even though this was frequently not related to the derivation of NOAELs relevant to risk assessment or whether additional safety factors were required. EFSA is not the lead EU agency for classification and labelling; this responsibility lies with the European Chemicals Agency (ECHA; previously with the ECB), therefore any decisions from the EFSA discussions would only be proposals rather than a final decision. There was a difference between some of the reports of the early EPCO meetings where tumours were discussed but classification was not a specific concern and later meetings where the classification proposal was critical. This reflects the move to the use of hazard trigger criteria for some aspects of the evaluation process, e.g. groundwater metabolites.
- **Interpretation of liver changes.** As mentioned elsewhere in this report, the interpretation of liver weight increases, hypertrophy and other histopathological changes is important to the risk assessments of many pesticides. For 29 active substances (13%) this was a key discussion point. Although there is some variation in the decision-making criteria, the general view was that liver weight increases up to *ca* 20% were not adverse *per se* but liver weight changes of > 10% accompanied by histopathology and / or clinical chemistry findings related to liver damage were regarded as adverse.
- **Body weight changes:** The interpretation of body-weight and body-weight-gain deficits was discussed for 11 active substances (5%). There was inconsistency in the use of absolute body weight or body weight gain as the parameter. There was general agreement that a change of >10% relative to controls in body weight gain was adverse. Reversible deficits seen at the start of dosing were sometimes treated as adverse and sometimes discounted.
- **Choice of extra safety factor / margin to LOAELs:** Discussions about the need for non-default (100) safety factors or whether adequate margins existed between reference doses and LOAELs or NOAELs for severe effects were frequent. It was not always clear from the descriptions why particular additional factors were chosen or the basis for the decision that a particular margin was acceptable.
- **Oral absorption:** The proportion of an oral dose that is absorbed is used in the derivation of a systemic AOEL (and potentially an AAOEL). For 46 active substances (21%) this was a major discussion point. The aspect most frequently mentioned was whether to include data from biliary cannulation studies when the liver was not the critical target organ for setting the AOEL.
- **Other aspects:**
    - duration of study applicable to AOELs for certain activities that might not be seasonal (e.g. seed treatments);
    - time of initial observation of effects of relevance to an ARfD (details often not in initial DAR);
    - bridging from data on related compounds (e.g. racemic mixtures to resolved isomer forms or from metabolites and major components);

- integration of multiple studies to give an overall (combined) NOAEL; evaluating multiple studies that give differing results – some took a precautionary approach, others concluded a non-reproducible finding was not adverse;
- for salts and esters, consistency of the expression of the dose relative to that used for the dietary and non-dietary exposures, and the need to make any corrections for relative molecular weights; need to correct dose levels in studies for low-purity test material as it was not always clear in study reports if the stated dose was corrected for purity or confirmed analytically;
- some DARs had limited quantitative information on critical aspects requiring the production of addenda or the checking of study reports during peer review meetings.

Options to reduce the need for discussions:

- Manufacturers must supply full impurity profiles for all toxicity batches and address any impurities present at significantly higher levels in the technical specification, in line with SANCO /10957/2003.
- Guidance to be developed on the use of HCD, covering use of generic databases, timeframes, format (e.g. mean, range, study by study) and appropriate end-points for HCD (e.g. clinical chemistry). This would enhance the information listed in the EU data requirements for pesticides.
- EFSA, ECHA and the Commission to agree a procedure for classification and labelling discussions for pesticide active substances to prevent duplication of effort.
- Guidance to be developed on the interpretation of liver weight and non-tumourigenic histopathology changes. Some texts on this are available (JMPR, 2006, USEPA) but if an EFSA (and ECHA?) agreed approach could be adopted, this should improve consistency of decision making.
- Guidance to be developed on the relevance of body weight changes. To include consideration of transient effects; relationship to food consumption; whether absolute weight or weight gain is the critical parameter and the significance of transient changes.
- Guidance on the use of extra factors for using LOAELs or for severe effects. Use of BMD analyses to support decisions on the magnitude of extra factors when using a LOAEL.
- Revise the current AOEL guidance (EU, 2006) to clarify oral absorption aspects and the duration of studies applicable to particular activities.
- DARs to include quantitative information on aspects of studies critical for determining NOAELs and to indicate if any effects possibly applicable to an ARfD are seen at the first measurements after the start of dosing.
- Guidance on bridging / read across. Considerable information already exists in the REACH guidance documentation and it could be readily transcribed to other areas where reference doses are set.

These proposals will not remove entirely the need for peer review discussions, since some chemicals will have complex toxicity profiles that need to be addressed by weight of evidence and expert judgement. However, the setting of scientifically-based criteria could reduce the need for discussions and should increase the consistency and transparency of the assessment process.

3. EFSA TASK 2:

Systematically and comprehensively evaluate EFSA's Conclusions on the ARfD (Acute Reference Dose) setting similarly to ADI and AOEL evaluations with an emphasis on exploring and gauging the usefulness of these evaluations in regard to the development of a concept for deriving a possible future "AAOEL"

The procedure adopted in analysing the ARfDs was essentially the same as that outlined above for the ADIs and AOELs.

3.1. Methodology

Copies of the EFSA conclusions on pesticide active substances were downloaded from the EFSA website[13]. The 224 EFSA conclusions on pesticide active substances published up to 1st December 2012 were evaluated to determine the basis for the derivation of the ARfD. These conclusions covered new active substances and those reviewed under the so-called List 2 and 3 procedures. The data for the following fields were compiled in Excel spreadsheets:

- date of the conclusion;
- value for the ARfD;
- NOAEL used to derive the ARfD and the associated LOAEL;
- value of the safety factor (SF) applied;
- reasons if the SF was not 100;
- end point(s) used to derive the ARfD;
- critical target organs or tissues;
- species used for the critical studies;
- particular attention was paid to end-points that might be secondary to local effects on the gastrointestinal tract;
- critical study type / duration;
- whether the original proposal in the DAR was supported;
- the ratio of the LOAEL to the NOAEL;
- additional comments.

The spreadsheets were used to perform analyses of these aspects:

- ranges of ARfDs, NOAELs and safety factors;
- common study types, species and durations used to derive the reference values;
- ratio of LOAELs to NOAELs;
- distribution of NOAELs associated with particular end-points;
- whether the original proposals by the rapporteur member states were supported.

In a number of instances it was necessary to go back to the Draft Assessment Report (DAR) to gather the required information; this applied to nearly all instances for the LOAEL value and original RMS proposals. The DARs are available on the EU website 'CIRCABC'[14] or on request from EFSA[15].

[13] http://www.efsa.europa.eu/en/publications.htm?scdtype=conclusion
[14] https://circabc.europa.eu/faces/jsp/extension/wai/navigation/container.jsp?FormPrincipal:_idcl=FormPrincipal:_id3&FormPrincipal_SUBMIT=1&id=8bd5dd33-9ab6-4b8c-a925-e1970255153a&javax.faces.ViewState=rO0ABXVyABNbTGphdmEubGFuZy5PYmplY3Q7kM5YnxBzKWwCAAB4cAAAAAN0AAE1cHQAKy9qc3AvZXh0ZW5zaW9uL3dhaS9uYXZpZ2F0aW9uL2NvbnRhaW5lci5qc3A=
[15] http://dar.efsa.europa.eu/dar-web/provision

In addition to the EFSA conclusions, analyses were performed on compounds for which both EFSA and the JMPR had set reference values (n=57). The JMPR summary information is available on its website[16], together with further details in the associated toxicological monographs[17]. The aim of this analysis was to determine what degree of consistency there was between the two organisations and the reasons for any differences.

The main discussion points identified during the written commenting phase and the 'EPCO' and 'PRAPeR' peer review meetings were evaluated to determine if there were any topics that routinely presented difficulties and should be prioritised when developing future guidance. The basic information in the form of commenting and evaluation tables and meeting reports is available from 'CIRCABC'[18].

The supporting data for the analyses summarised below are presented in a number of spreadsheets attached to the report as searchable Excel files. The contents of the spreadsheets are explained in Annex 1.

The overall spreadsheets also contain some information on the so-called List 4 review compounds. Initially, these compounds were included in the summary analyses but they have now been removed. The List 4 compounds typically have limited databases, because they are naturally occurring or their main uses are not as pesticides; therefore, the derivation of the ARfD often requires approaches that are not routinely used for active substances from other lists/groupings and the patterns of results are inconsistent with those of synthetic, chemical active substances.

Additional evaluations related to the ARfD as a potential basis for setting an AAOEL were performed. These were comparisons of oral and dermal LD50 values for anti-cholinesterase compounds that would be expected to have high peak concentration (Cmax)-type-effects and comparisons of dermal and oral developmental toxicity results.

Rationale for conclusions

The rationale for the choice of individual NOAELs and ARfDs was not always clearly described in the documentation. This was particularly so when the standard or default approaches (e.g. as described in WHO, 2009), with statistically significant findings treated as adverse and a SF of 100, were used. More details of the underlying rationales were normally, but not always, provided when non-default SFs were used, findings were set aside as non-relevant to humans or combined NOAELs were derived from two or more studies. Information on the rationales is provided in the spreadsheets and key aspects are highlighted in the summary texts.

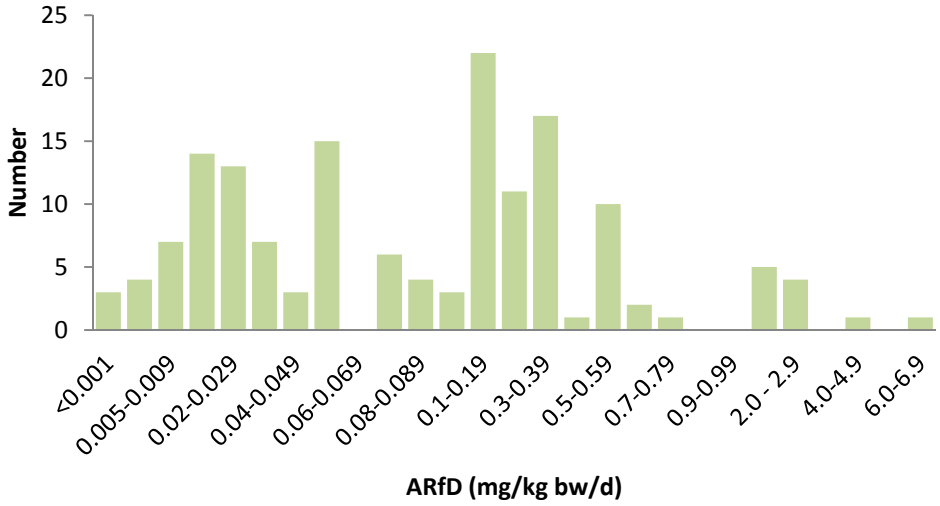
3.2. Conclusions on acute reference doses

ARfD (n= 224; values set for 154)

[16] http://www.who.int/entity/foodsafety/chem/jmpr/publications/pesticide_inventory_report_2010.pdf
[17] http://www.who.int/foodsafety/chem/jmpr/publications/monographs/en/index.html
[18] https://circabc.europa.eu/faces/jsp/extension/wai/navigation/container.jsp?FormPrincipal:_idcl=FormPrincipal:left-menu-link-lib&FormPrincipal_SUBMIT=1&javax.faces.ViewState=rO0ABXVyABNbTGphdmEubGFuZy5PYmplY3Q7kM5YnxBzKWwCAAB4cAAAAAN0AAE4cHQAKy9qc3AvZXh0ZW5zaW9uL3dhaS9uYXZpZ2F0aW9uL2NvbnRhaW5lci5qc3A=

Range of values (mg/kg bw/d)

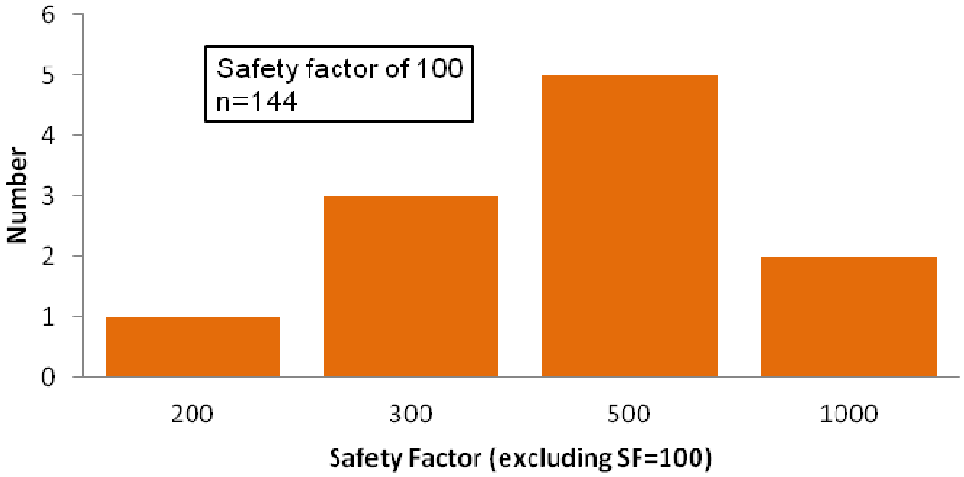
0.00015 – 6.3 Mean 0.22 Median 0.08



Safety factor not 100

N= 11

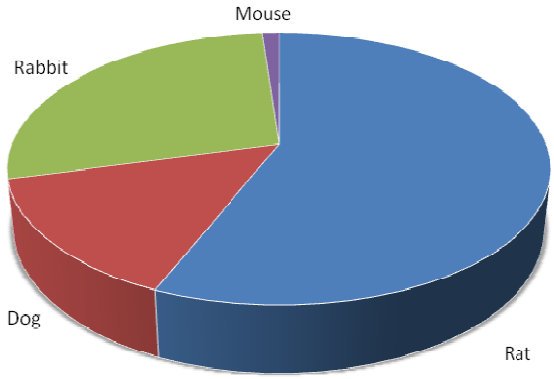
Range 200 - 1000



Main Reasons:

- Provide margin of safety to severe effects 2
- Use of LOAEL 3
- Severity / tumours 5

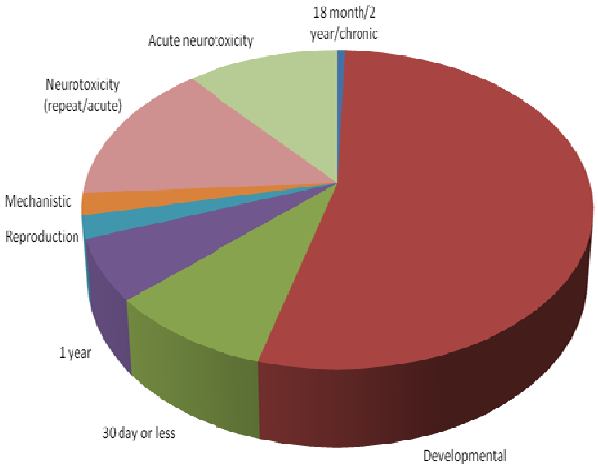
3.1.2. Frequency of species used to set ARfD



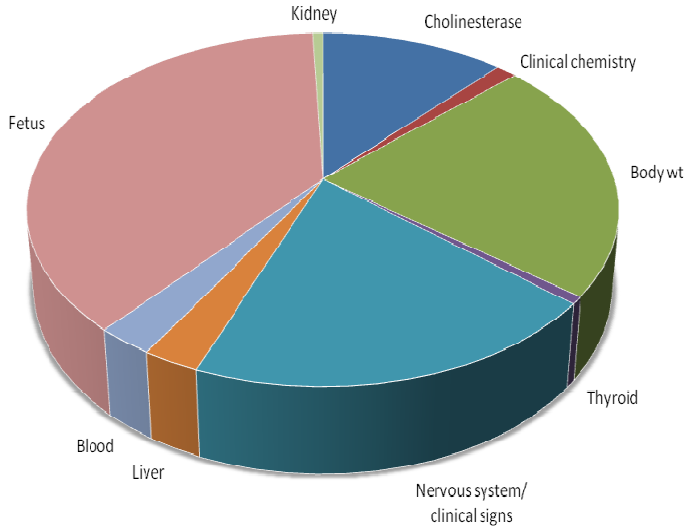
Rat	58%
Dog	16%
Mouse	1%
Rabbit	29%
Human	0

Frequency of main study types used to set ARfD

18 month / 2 year / chronic	1%
1 year	8%
90 day	8%
2 - 30 day	10%
Mechanistic	3%
Developmental	58%
Reproduction	3%
Neurotoxicity (repeat /acute)	19%
Acute neurotoxicity	16%



Main targets / end-points



Liver	3%
Thyroid	1%
Blood	3%
Cholinesterase	11%
Clinical chemistry	1%
Body wt	21%
Nervous system / clinical signs	23%
Fetus	34%
Kidney	1%

No value set

N= 54

Main reasons

- Inadequate data / missing information 2
- No exposure / within background 2
- No acute alerts 50

Analysis of LOAEL to NOAEL ratio

For ARfDs set by EFSA, the ratio of LOAEL to NOAEL for the critical study(ies) varied from 1.25 to 30. For 45 compounds the ratio was 2 or less and only 4 compounds had ratios in excess of 10. The mean ratio was 3.3, and the median ratio was 3. These results are less variable than for ADIs and AOELs and indicate that dose spacing in most short-duration studies is reasonably well optimised.

Analysis of original DAR proposals

For 115 compounds (50%), the original proposal by the RMS was in agreement with the final ARfD value. Reasons for changing the proposal included different safety factors, new data submitted during the procedure, alternative study duration, and a different NOAEL chosen.

Comparison with values derived by JMPR

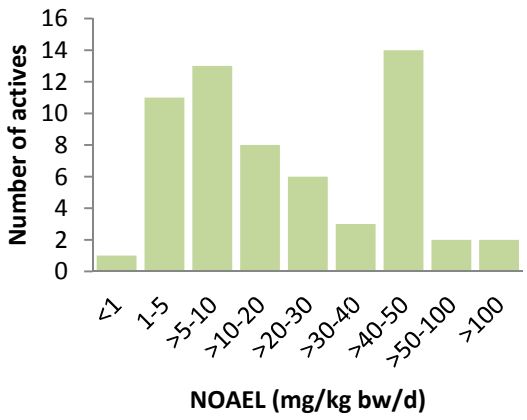
Fifty-seven active substances had been considered by both JMPR and EFSA for the derivation of ARfDs. A comparison of the values is summarised in Table 2. For 24 compounds (42%) both groups had the same conclusion. JMPR was more likely than EFSA to conclude that the setting of an ARfD was not necessary (7 versus 1). JMPR was also more likely than was EFSA to base its conclusions on data derived from human studies, use a CSAF, and to set separate values for women of child-bearing age. Overall, where there was a difference in the ARfD value, those concluded by JMPR were normally higher than those of EFSA (70% of compounds where decisions differed). The level of concordance for ARfD setting is lower than for ADI setting; this appears to be due to a combination of factors including methodological differences such as the use of human data and CSAFs and the fact that the procedures for deriving ARfDs were developing during the 1990s when a number of the JMPR assessments were performed, whereas all the EFSA derivations were post-2002.

Table 2 – Comparison of JMPR & EFSA ARfDs

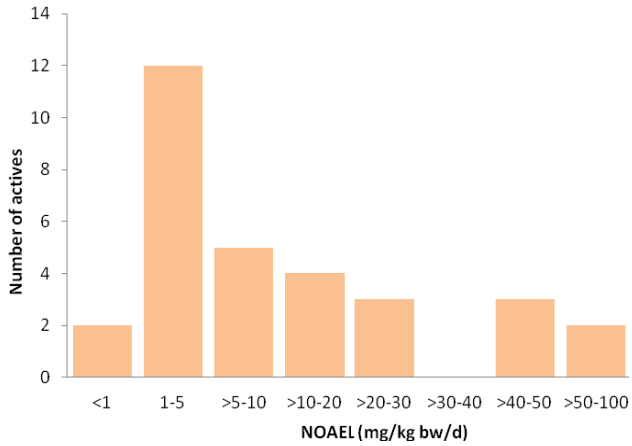
Criterion	Number
Both have same value	9
Values differ only due to rounding	4
Agree ARfD not required	11
JMPR value higher than EFSA	23
EFSA value higher than JMPR	2
Only JMPR conclude not required	7
Only EFSA conclude not required	1
Total	**57**

Analysis of NOAELs for particular end-points used for setting ARfDs

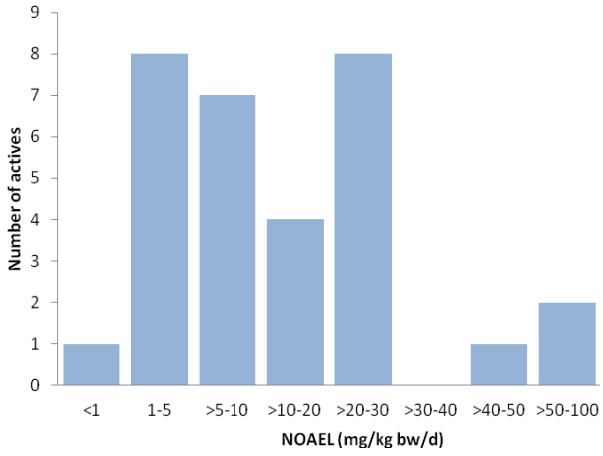
ARfd Foetus



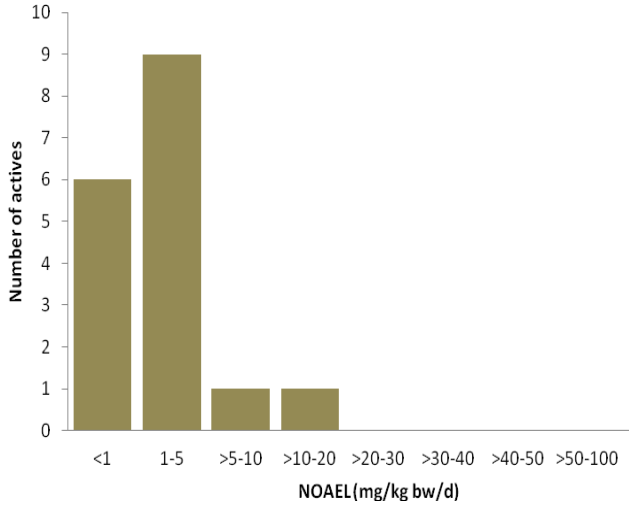
ARfD nerve and sign



ARfD bd wt



ARfD cholinesterases



These plots of the NOAELs for fetal effects, reduced body weight, clinical signs and neurological effects (including motor activity) indicate that other than for cholinesterase inhibition, there is a broad range of potencies and none of the end points can be considered as being highly sensitive.

3.3. Considerations for AAOEL derivation

Evaluation of developmental toxicity studies using the dermal route of exposure.

The production of malformations in developmental toxicity studies is generally considered to be a potentially acute effect, since they occur when exposures take place during a critical time window. The effects are considered to be related to Cmax rather than AUC. Therefore an analysis of dermal developmental toxicity studies conducted with some pesticide active substances was performed to see if malformations seen following oral exposure are induced following dermal exposures. It is accepted that the database is likely to be somewhat unrepresentative, as for most compounds a dermal developmental study is only performed if marked effects have been found via the oral route. However,

the database is considered adequate to investigate the potential for severe fetal effects to be produced via the dermal route. The data are summarised in Table 3 below.

The data show that for binapacryl, spiroxamine, epoxiconazole, dinoseb, dinocap, vinclozolin and triapenthol, malformations were seen following dermal exposure at dose levels not markedly different from those producing related findings orally. For flusilazole, although there were no malformations, developmental effects were seen following dermal exposure at lower dose levels than the oral NOAEL. Of particular note was the finding that for a number of the active substances considered, reductions in body weight gain were noted following dermal exposures.

No attempt has been made to relate these findings to available dermal absorption data as the dermal developmental studies were generally performed with relatively high doses of the active substance in aqueous vehicles that differ significantly from the formulated products tested for dermal absorption determination.

Table 3 – Summary of developmental studies conducted by the dermal route

Compound	**Species**	**Author**	**CRD DPDB ref**	**Malformations* dermally**	**Other effects dermally**	**Oral malformations**
Binapacryl	Rabbit	Becker et al	412	At 150 mg/kg bw/d and above	Post implantation loss at 150 mkd	
Cyanazine	Rabbit	Rose	434	No at 2000 mkd	Significant maternal toxicity	
Fentin hydroxide	Rabbit	Doherty	21635	Equivocal increase at 3 mkd	Reduced maternal body wt	
Spiroxamine	Rat	Becker & Biedermann	23162	Cleft palate 100 mkd	Local effects at 5 mkd	Cleft palate 100 mkd
Bronopol	Rat	James & Palmer	30733	No at up to 250 mkd	no	
Isazophos	Rabbit	Sabol	30950	No at up to 150 mkd	Deaths (13/18 at 150 mkd)	
Fenoxaprop-ethyl	Rat	Becker	33026	No at up to 1000 mkd	No at up to 1000 mkd	At 100 mkd
Fenoxaprop-ethyl	Rabbit	Becker	33148	No at up to 1000 mkd	No at up to 1000 mkd	At 50 & 200 mkd
epoxiconazole	Rat	Hellwig	39176	1 cleft palate at 1000 mkd	Maternal toxicity	Lots of cleft palate 180 mkd
Tebuconazole	Rat	Renhof	41235	No at 1000 mkd	None	At 100 mkd
Dinoseb	Rabbit	Becker et al	60886	Yes at 30 mkd		
Dinoseb acetate	Rabbit	Becker et al	60901	Yes at 30 mkd	Local irritation at 10 mkd	Microphthalmia ca 15 mkd in rat
Dinocap	Rabbit	Costlow & Lutz	62039	No at up to 100 mkd	NOAEL 50 mkd	Yes at 3 mkd
Dinocap	Mouse	Costlow & Lutz	62073	Cleft palate & otoliths at 25 mkd		Cleft palate at 10 mkd
Vinclozolin	Rat	Hellwig	63784	↓anogenital distance 180 mkd		↓anogenital distance 50 mkd
Methiocarb	Rabbit	Biedermann & Dotti	74444	No at 250	General toxicity at 50	No
Bromuconazole	Rat	Higgins	75516	No at up to 400 mkd		At 70 mkd

Compound	Species	Author	CRD DPDB ref	Malformations* dermally	Other effects dermally	Oral malformations
Triapenthenol	Rat	Becker et al	98146	Extra ribs at 100 mkd	no	
Triapenthenol	Mouse	Becker et al	98147	Cleft palate at 100 mkd	no	
Flusilazole	Rat	Schardein	102229	No	↓placental size, developmental effects at 10 mkd	No, but developmental NOAEL 50 mkd
Acibenzolar-s-methyl	Rat	Khalil	47756	No at up to 500 mkd	No	No at 350 mkd

* True malformations, not just anomalies and variants

Comparison of rat dermal and oral LD50 values for anticholinesterase active substances and products

A review of data in DARs and other summary documents for oral and dermal LD50 values for 22 anticholinesterase active substances and products was performed to obtain an impression of whether acute neurotoxic effects seen following oral exposure are induced following dermal exposure. Anticholinesterase compounds were chosen because inhibition of acetylcholinesterase is a well-understood Cmax-dependent effect (especially for carbamates) that could lead to lethality at high doses. LD50 data were chosen as these are one of the few study types where comparable oral and dermal data are available on a good number of compounds / products. The results are summarised in Table 4 below. In many cases several values were available and these have been presented as ranges or approximate (*ca*) values. In most cases, dermal LD50 values were significantly higher than those for the oral route. This shows that in the majority of examples, exposure via the dermal route significantly diminishes the acute toxicity, as determined by the LD50. There are several caveats to this analysis; in particular, it is not applicable tor non-lethal single-dose effects that might be adverse for a short period of time, but reversible; and, in a number of cases the dermal exposure will not all be in contact with the skin but will be present as a layer of finite thickness.

Exceptions to the broad conclusion of much reduced toxicity following dermal exposure applied to the following substances:

- Aldicarb, using certain vehicles;
- Dichlorvos and a 55% emulsifiable concentrate formulation;
- Parathion under certain conditions;
- Cadusafos where the rabbit dermal LD50 is similar to the rat oral LD50. No directly comparable rabbit oral data were located;
- For ethoprophos, a rabbit dermal LD50 of 8 – 9 mg/kg bw was cited, compared with an oral developmental study in rabbits using repeat doses of ≥10 mg/kg bw/day. However, ethoprophos is a severe irritant.

Table 4 – A comparison of LD50 values obtained from oral and dermal routes of administration in anticholinesterase active substances and products

Compound	Rat oral LD50 (mg/kg bw)	Rat dermal LD50 (mg/kg bw)
Aldicarb active	3 – 40	*Ca* 1
Aldicarb 10% granule	*635 – 3200*	*8*

Compound	Rat oral LD50 (mg/kg bw)	Rat dermal LD50 (mg/kg bw)
Bendiocarb	560	*Ca* 100
Benfuracarb	2000	100 - 200
Cadusafos active	Rabbit 11 – 40	Rat 30 - 80
Cadusafos 20% CS	*>5000*	*1097*
Carbaryl	>5000	600
Carbofuran	1000	7
Carbosulfan	3700	*Ca* 100
Chlorpyrifos active	2000	60 - 200
Chlorpyrifos 48% EC	*2000 – 5000*	*400 -500*
Diazinon active	>2000	*Ca* 1000
Diazinon 60% EC	*>2000*	*Ca 1000*
Dichlorvos active	120 – 260	60 - 80
Dichlorvos 55% EC	*30 – 50*	*30 - 50*
Dimethoate active	>2000	300 - 400
Dimethoate 40% EC	*>2000*	*300 - 400*
Ethoprophos	200 - 1200	40 - 80
Methiocarb active	>5000	30 – 50
Methiocarb 50% FS	*>5000*	*25 - 50*
Methomyl active	>2000	20 - 30
Methomyl 20% granule	*>5000*	*132*
Oxamyl active	2000	*Ca* 3
Oxamyl 10% granule	*>2000*	*Ca 40*
Parathion active	4 – 100	2 - 10
Parathion 50% EC	*135*	*1 - 10*
Phosalone active	1530	120 - 165
Phosalone 35% EC	*>2000*	*>2000*
Pirimicarb active	>2000	140 – 150
Pirimicarb 50% granule	*>2000*	*50 - 100*
Pirimiphos-methyl active	>2000	1414
Pirimiphos-methyl 50% EC	*>2000*	*1500*
Thiobencarb active	>2000	*Ca* 1000
Thiobencarb 90% EC	*>2000*	*Ca 1000*
Thiodicarb active	>2000	50 – 200
Thiodicarb 37% SC	*>2000*	*386*
Triazamate active	>5000	50 -200

Compound	Rat oral LD50 (mg/kg bw)	Rat dermal LD50 (mg/kg bw)
Triazamate 14% EW	*>2000*	*386*
Trichlorfon active	>5000	200 - 260
Trichlorfon 80%	*>2000*	*395 - 993*

Conclusion

These simplistic evaluations of dermal LD50 and dermal developmental toxicity studies show that in some instances severe effects (death or malformations) that are considered to be Cmax-dependent can be produced by dermal exposures. Therefore it is reasonable as a default assumption that any compound with an ARfD will require an acute non-dietary assessment. The initial basis for deriving an AAOEL (surrogate AAOEL) for this non-dietary assessment could be the ARfD corrected for oral absorption. The oral absorption correction should not automatically be taken from that used for the AOEL if there is a significant biliary contribution as the end-points for the AAOEL and AOEL might differ.

If the exposure estimate is above the surrogate AAOEL, a number of options are available to define or refine an AAOEL based on an ARfD:

- the ARfD can be refined with an oral study using the approach outlined in Solecki et al (2005);
- kinetic data could be used to demonstrate that effects seen in oral studies (especially gavage) that are Cmax-related would not occur following dermal exposure. This data would need to include some consideration of the relevant pesticide formulations, not just the active substance in a simple vehicle;
- a route-specific study can be performed to investigate effects relevant to dermal and inhalation exposures.

It is considered that there would be no value in performing a routine inhalation study (according to the OECD test guideline) for active substances that are not highly volatile or applied in a manner likely to generate a significant proportion of particles that would be respirable and reach the alveoli. Regulatory tests require the generation of atmospheres containing particles with a mass median aerodynamic diameter of *ca 3* μm. Such atmospheres are unrepresentative of exposures from pesticide application, which will be typically to particles of > 50 μm (see Table 5). Particles of >50 μm will not reach the alveoli but will be caught in the upper respiratory tract and absorbed via the gastrointestinal tract following mucociliary clearance. In addition to the particle size aspect, the predominant route of exposure from pesticides is dermal. Assuming 100% inhalation absorption and 15% dermal absorption, over 90% of the systemic dose for operators and bystanders comes from the dermal route for a range of scenarios that include airblast application in orchards (P. Hamey pers comm. Jan 2013). Therefore any route-specific data for the derivation of an AAOEL should be generated via the dermal route. Any studies should use a vehicle or blank formulation that would maximise dermal absorption or be representative of those likely to be used in pesticide products.

Table 5 – Droplet size from typical spray nozzles

Table 5a - Measurements made with Malvern SprayTec (spatial sampling):

Nozzle	Classification boundary	%<50 um	%<100 um	VMD, um
Bcpc 01	Fine-very fine	6.41	32.49	123
Bcpc 03	Fine - medium	2.4	14.43	174
Bcpc 06	Medium - coarse	1.42	8.78	229
Bcpc 88	Course-very coarse	1.03	6.46	314
Bcpc 10	Very coarse – extra coarse	0.65	4.68	360

Table 5b - Measurements made with Oxford Lasers Visisizers for the fine-medium boundary nozzle at three magnifications (temporal sampling):

Nozzle	Magnific-ation	Min droplet size, um	Classification boundary	%<50 um	%<100 um	VMD, um
Bcpc 03	X 2	16.9	Fine - medium	1.31	6.68	229
Bcpc 03	X 1	33.5	Fine - medium	0.67	6.64	247
Bcpc 03	X 0.58	56.8	Fine - medium	0	3.4	245

A comparison of the relative exposures based on 75 centile values, as used for repeat exposure assessments, and 95 centile values, as proposed for acute assessments, showed that the difference will be a factor of approximately 6 (P. Hamey pers comm. Jan 2013). Some exposure assessments for repeat exposures give values close to the AOEL; some AOELs are the same or similar to the ARfD for the active substance. Therefore it is likely that specific AAOELs will be required for a moderate number of active substances.

3.4. Consideration of local effects

Many pesticide active substances induce local effects (i.e. effects at the initial site of contact) in addition to systemic effects. Whilst systemic effects are covered by the reference values currently derived for pesticides (ADI, AOEL and ARfD), local effects are not specifically taken into account by these systemic standards (even though they might still be partly covered).

It is general regulatory practice to consider further (in the assessment of a substance) only those local effects that lead to the classification (under the Dangerous Preparations Directive or the CLP Regulation) of the product (containing the substance) for irritation/corrosivity (on skin, eye or respiratory tract) or sensitisation (on skin or respiratory tract). Other local effects that do not lead to the classification of the product are not considered severe enough to require further consideration.

Guidance on how to assess the risks posed by these local effects is being developed within the framework of the Biocidal Product Regulation. A qualitative approach is being proposed, which relies on risk management measures (RMMs), including personal protective equipment (PPE): the more severe the hazard classification for these local effects, the more stringent the RMMs required to ensure risks are acceptable. Where PPE cannot be relied upon (e.g. non-

professional users, bystanders, residents), then other risk management measures should be considered.

Only in exceptional circumstances (e.g. when reliable quantitative dose-response data that are relevant to the substance in the product are available) can a more (semi)quantitative approach be followed. This requires the derivation of an inhalation acceptable exposure concentration (AEC) expressed in mg/m^3 for respiratory effects or a dermal AEC expressed as a percentage concentration or a dose per unit area (mg/m^2) for skin irritative or sensitising effects. For local effects on the gastro-intestinal tract, usually an oral AEC does not need to be derived because these effects are generally covered by the systemic reference values (as these are usually derived from NOAELs identified from oral studies and such oral NOAELs are established on the basis of local and systemic adverse effects).

As also indicated in the BRAWG (Bystander Risk Assessment Working Group) (2012) report[19], fully quantitative dose-response information for skin sensitisation is rarely available from standard animal studies. This makes any quantitative risk assessment approach for skin sensitisation a challenge. It is also noted that current exposure models are not suitable for the estimation of localised peak exposure levels on the skin (e.g. amount per unit area) which are the determinants of both dermal irritation and dermal sensitisation. This adds to the challenge of performing a fully quantitative risk assessment for dermal local effects.

The BRAWG report concludes that if bystanders and residents are exposed to dilutions of a product, which still trigger classification for local effects, then a risk of effects may still be present and risk managers should be informed.

3.5. Consideration of a chronic AOEL

In addition to the Acute AOEL (AAOEL) the issue of chronic exposures for residents has been raised. Resident exposures to pesticides will generally decline over time as the active substance volatilises or degrades in the environment. However, some pesticides are applied several times throughout the year potentially resulting in exposures over a prolonged period. Depending on the persistence of the active substance it might be that a comparison with a long-term or chronic-AOEL is required. In such instances there is no reason why, as a first tier, the risk assessment can not be performed by comparing the exposure with a systemic, chronic-AOEL derived by taking the ADI and correcting it for the extent of oral absorption. In determining the relevance of a chronic AOEL based on an oral study the primary route of exposure would need to characterised and considered on a case-by-case basis. Refinement might be possible (e.g. if the ADI is based on gastrointestinal effects) but this would need to be justified.

[19] http://cot.food.gov.uk/pdfs/brawgreport.pdf

3.6. Summary

The analysis of the 224 pesticide active substances with EFSA conclusions and ARfD considerations produced results with different patterns to those for the ADI and AOEL but broadly in line with the existing guidance documents for ARfD derivation. Some of the end-points routinely used to derive ARfDs could be related to gastrointestinal effects (e.g. body weight effects secondary to reduced food consumption (N=33)) or local irritative effects (n=4) and might not be directly applicable to AAOEL derivation where dermal or respiratory exposure routes predominate.

- The majority of ARfDs were based on rat studies, with rabbit developmental studies used in approximately one quarter of cases.
- Developmental studies were used in *ca* half the ARfDs derived. The relevance of gavage developmental studies to dermal exposure is a debatable point but cannot be automatically discounted.
- The commonest target was the fetus (various effects), used in approximately half the cases.
- Body weight changes, often related to reduced food consumption, and clinical signs (especially in dogs) were used in 25% of cases.
- The default safety factor of 100 was used in the majority of cases (*ca* 90%).
- Acute toxicity studies (e.g. acute neurotoxicity) were used in only 20 instances (*ca* 13%) indicating a significant number of ARfDs are likely to be conservative as they are based on repeat-dose effects.
- Only a small number of ARfDs were based on pathological effects on major organs such as the liver, kidney or heart. This is contrary to the results obtained from the ADI and AOEL analyses.
- The primary reason for not setting an ARfD is the absence of acute alerts. It was not always clear how extensive the consideration of potential acute alerts was. For example, for some compounds, blood effects were used for deriving ARfDs yet for others, blood effects were present in repeat-dose studies but no ARfD was set.
- JMPR was more likely than EFSA to conclude that an ARfD was not necessary. When there were discrepancies between the values set by EFSA and JMPR, the main reason was the choice of study from which to choose a NOAEL; the values set by JMPR tended to be higher than those set by EFSA.
- The range of values in this sample is consistent with that of the broader EU database (0.0015 – 10 mg/kg bw/day[20]) and the findings were in line with those of Solecki et al (2010) who reviewed ARfDs set in the EU between 2000 & 2008.
- Analyses of dermal acute toxicity data and dermal developmental toxicity data shows that there are instances when effects relevant to acute exposures can be see following dermal exposure. Therefore the default assumption must be that an AAOEL should be considered in all instances and certainly when an ARfD has been set.

[20] http://ec.europa.eu/sanco_pesticides/public/index.cfm?event=activesubstance.selection

4. EFSA TASK 3:

Comprehensively collect and systematically evaluate relevant scientific literature or other relevant publications or information sources (like guidance documents or government reports) related to the principles of identifying and characterising the critical reference points for hazard and risk assessments outside the EU and outside pesticide evaluations that could be relevant for future ADI, AOEL and "AAOEL" settings for pesticides in the EU (e.g. derivation of TDIs)

4.1. Results of the literature review.

Literature searches were performed by information scientists using the search criteria described in Annex 5. Based on the abstracts obtained, potentially relevant references were identified by the toxicology specialists, obtained and evaluated. The toxicologists also performed targeted online searches and obtained copies of references cited in core texts and primary references.

In addition a questionnaire was distributed to EU and non-EU agencies involved in human health risk assessments of chemicals. The questionnaire asked about approaches to the derivation of health based guidance values and sought details of available guidance documents.

The main reference values applied to the assessment of chemicals for human health effects as required under some different EU regulatory schemes or recommended by different organisations are briefly described below.

The *Tolerable Daily Intake* (*TDI*) is an estimate of the amount of a substance in food and drinking-water, expressed on a body weight basis (mg/kg or µg/kg of body weight), that can be ingested over a lifetime without appreciable health risk (WHO, 2011). The TDI is essentially the same as the ADI, but, for chemical contaminants, which usually have no intended function in drinking-water, the term 'tolerable daily intake' is deemed to be more appropriate than 'acceptable daily intake,' as it signifies permissibility rather than acceptability. Wherever possible, the NOAEL / LOAEL is based on long-term studies, preferably of ingestion in drinking-water. However, N(L)OAELs obtained from short-term studies and studies using other sources of exposure (e.g., food, air) may also be used. Over many years, JECFA and JMPR have developed certain principles in the derivation of ADIs. These principles have been adopted where appropriate in the derivation of TDIs used in the development of guideline values for drinking-water quality. The application of uncertainty factors follows the same principles as that for the derivation of ADIs. For contaminants for which there is sufficient confidence in the database, a smaller uncertainty factor can often be applied. For most contaminants, however, there is greater scientific uncertainty, and a relatively large uncertainty factor is used. Since drinking-water is not usually the sole source of human exposure to the substances for which guideline values have been set, guideline values derived with the TDI approach take into account exposures from all sources by apportioning a percentage of the TDI to drinking-water, so that total daily intake from all sources does not exceed the TDI.

The WHO (2011) considered the use of *categorical regression* (IPCS, 1994) as an alternative to the TDI. In this approach, data on toxicity are classified into one of several categories, such as no-observed-effect level (NOEL) or NOAEL, or others, as appropriate. These categories are then regressed on the basis of dose and, if required, duration of exposure. The result is a graph of probability of a given category of effect with dose or concentration, which can be useful in the analysis of potential risks above the tolerable intake, especially for comparisons among chemicals. Categorical regression utilises information from the entire dose–response curve, resulting in more precise estimates of risk when compared with the current TDI. However, the WHO concluded that

categorical regression requires more information and the interpretation of the probability scale can be problematic.

The WHO (2011) has also considered ways to derive guideline values for non-threshold effects of contaminants in drinking-water. For such genotoxic carcinogens, guideline values are normally determined from a hypothetical mathematical model. The guideline values reported were the concentrations in drinking-water associated with an estimated upper-bound excess lifetime cancer risk of 10^{-5} (or one additional cancer per 100,000 of the population ingesting drinking-water containing the substance at the guideline value for 70 years). It was noted that the models used are presumed to be conservative.

EFSA's Panel on Contaminants in the Food Chain (CONTAM) has defined health-based guidance values for chemical contaminants in food and feed (Alexander *et al.*, 2012). The preferred point of departure from animals studies is the BMDL (see task 4); otherwise, a suitable NOAEL is selected. Assessment factors are then applied. In some cases, the CONTAM Panel has been able to model human data and to incorporate information from biomarkers of exposure or of effect in the characterisation of the hazard. This allows the use of a *body burden approach*, where an estimate of systemic exposure (body burden), rather than external dose, is used in the risk characterisation. Usually, however, guidance values are derived in a similar manner to those for pesticide active substances. An *ARfD* is established for those chemical contaminants that could give rise to acute health effects in relation to short periods of intake. Conversely, when a substance shows a long biological half-life, tends to accumulate in the human body and exposure over a longer time period therefore matters, the CONTAM Panel usually establishes a *tolerable weekly intake* (*TWI*) as the guidance value.

EFSA and the Joint FAO/WHO Expert Committee on Food Additives (JECFA) independently derived guidance values for cadmium in food (EFSA, 2009a; FAO/WHO, 2011). The FAO/WHO established a *provisional tolerable monthly intake* (*PTMI*) for cadmium which, when converted to a weekly intake, differed from the TWI established by EFSA. Both assessments used the same epidemiological dataset and had two primary components, a concentration-effect model that related the concentration of cadmium in urine to that of a biomarker of renal tubular effects, and a toxicokinetic model that related urinary cadmium concentration to dietary cadmium intake. Methodological differences were identified to account for the different values obtained: the identification of the starting point on the basis of the concentration-effect model (EFSA used a hybrid BMD approach, whereas JECFA used a linear model fitted to the data); the statistical approach to account for the variability and uncertainty in the biomarkers; and the methodology for transforming urinary cadmium concentrations into dietary intakes (EFSA Panel on Contaminants in the Food Chain (CONTAM), 2011a).

Many substances that the CONTAM Panel assesses are non-threshold genotoxic carcinogens. Until 2005, the advice given by the risk assessor to the risk manager was to reduce exposure to such substances to a level that is as low as reasonably achievable (known as the ALARA principle). However, such advice did not provide risk managers with a basis for setting priorities for action, either with regard to the urgency or to the extent of measures that may be necessary. To overcome this, the EFSA Scientific Committee proposed the margin of exposure (MoE) approach (see task 6) as a harmonised approach for the risk assessment of substances that are both genotoxic and carcinogenic. The MoE approach is not confined to such substances, however: it has also been applied to cases where the data are insufficient or otherwise considered inappropriate to establish a health-based guidance value. For example, the CONTAM Panel considered it appropriate to calculate MoEs to support the risk characterisation of lead (EFSA Panel on Contaminants in the Food Chain (CONTAM), 2010). The CONTAM Panel identified developmental neurotoxicity in young children and cardiovascular effects and nephrotoxicity in adults as the critical effects for the risk assessment.

The Panel then calculated respective BMDLs for these effects from blood lead levels, which were then extrapolated to external exposure levels for comparison to estimated dietary exposure in various human population subgroups.

Sometimes, the available data on a food contaminant do not allow either the establishment of a health-based guidance value or the calculation of a BMDL for use as a starting point in the MoE method. In this case, the 'threshold of toxicological concern (TTC) approach' (see task 4) might be applied (for example, EFSA Panel on Contaminants in the Food Chain (CONTAM), 2011b).

The assessment of animal health risks associated with the presence of chemical contaminants in animal feed, and subsequently potential risks to human health, follows the same principles as the human health risk assessment. However, in the hazard characterisation, species-specific and inter-species differences in animals need to be taken into account. The exposure assessment and risk characterisation are based on the respective animal species and their specific diets. The hazard characterisation aims to identify the most relevant toxicological endpoint for the respective animal species to derive a safe intake level. Most often a NOAEL / LOAEL is identified, but a BMDL can also be used as a starting point (EFSA Panel on Contaminants in the Food Chain (CONTAM), 2011c).

The Food and Agriculture Organization and World Health Organization (JMPR, JECFA, IPCS) have also jointly produced guidance on health-based guidance values for chemicals in food (FAO/WHO, 2009), derived for either acute or chronic oral exposure. Several guidance values are described. Many of these (for example the ADI for substances intentionally added to food and residues; and the ARfD for acute effects or exposures of 24 hours' duration or less, mainly used for pesticide residues) and the means of deriving them are broadly in line with the EU / EFSA considerations. For unavoidable food contaminants, JECFA calculates tolerated intakes. It has established the concept of a *provisional tolerable weekly intake* (*PTWI*) for contaminants with cumulative properties. The use of the term 'provisional' expresses the tentative nature of the evaluation, in view of the paucity of reliable data on the consequences of human exposure at levels approaching those with which JECFA is concerned. *Provisional maximum tolerable daily intake* values (*PMTDIs*) are established for food contaminants that are known not to accumulate in the body. For contaminants that may accumulate within the body over a period of time, JECFA has used the *PTWI* and *provisional tolerable monthly intakes* (*PTMI*). On any particular day, consumption of food containing above-average levels of the contaminant may exceed the proportionate share of its weekly or monthly tolerable intake (TI). JECFA's assessment takes into account such daily variations, its real concern being prolonged exposure to the contaminant, because of its ability to accumulate within the body over a period of time. JECFA also considers that, if several substances that produce similar toxic effects are to be considered for use as food additives, pesticides or veterinary drugs, or may appear as contaminants, it might be appropriate to set group ADIs or TIs.

The point of departure in the derivation of guidance values by JECFA is generally the lowest NOAEL, LOAEL or BMDL in the most sensitive species. However, when relevant, JMPR and JECFA use an overall NOAEL as a basis for the ADI, considering the most relevant studies together. JMPR noted that there is often available more than one study in which the same end-points have been addressed. In such situations, the dose spacing may be different, resulting in different NOAELs and LOAELs. It was agreed that in such circumstances it might be appropriate to consider the studies together. When they are comparable, including consideration of study design, end-points addressed, and strain of animal, the 'overall NOAEL' would be the highest value identified in the available studies that provides a reasonable margin ($\geq$ 2) over the lowest LOAEL, provided that due consideration is given to the shape of the dose–response curve (FAO/WHO, 2004). For dose-response modelling, JECFA does not consider it necessary to model each observed end-point in each study; rather, end-points are selected as candidates for modelling based on the toxicological impact together with the apparent magnitude of the response, and the suitability of the data for modelling. If suitable

information is available, chemical specific adjustment factors (CSAF, see task 6) can replace the default 100-fold factor (FAO/WHO, 2009).

For its risk assessments, the USEPA has produced guidelines for each of the major human health end-points[21] (published between 1986 and 2005). For developmental toxicity, the 'reference dose for developmental toxicity - RfD_{DT},' based on short-term exposure is derived (the subscript $_{DT}$ is added to distinguish it from the reference dose (RfD), which is used for chronic (lifetime) exposure situations). In the assessment of carcinogens, linear extrapolation is applied in cases where the mode of action is mutagenic, linear but non-mutagenic or cannot be determined. For non-linear modes of action, the default reference value is the RfD or RfC. To provide additional protection for early-life exposures, age-dependent adjustment factors are applied in the absence of chemical-specific data to evaluate differences in the responses of adults and juveniles. The RfD_{DT}, RfD and RfC values are calculated from a point of departure (NOAEL, LOAEL, BMD) divided by uncertainty values for interspecies differences in response, intraspecies variability and deficiencies in the database. The USEPA has recently published technical guidance on the use of the BMD approach to aid consistency in the use of this tool in its risk assessments (USEPA, 2012a).

When whole-mixture data or data on a sufficiently similar mixture are not available for dioxin-like compound (DLC) exposures, the USEPA (2010) recommends the use of the consensus mammalian *toxicity equivalency factors* (*TEF*) values from WHO (van den Berg et al., 2006) in the assessment of human health risks posed by exposures to mixtures of tetrachlorinated dibenzo-*p*-dioxin (TCDD) and DLCs, with TCDD as the index chemical. The TEF methodology is based on the concept of dose addition (USEPA, 2000). With this method, the combined toxicity of the individual components is estimated from the sum of their doses, which are scaled for potency relative to that of another component of the mixture for which adequate dose-response information is available. In practice, the scaling factor for each DLC is typically based on a comparison of its toxic potency to that of a designated index chemical. The index chemical is well-studied toxicologically and must have a dose-response function to apply the methodology to an environmental mixture. A comparative measure from an individual toxicity assay is termed an estimate of relative potency (ReP). Based on the RePs that may be estimated from multiple toxicological assays, each individual PCDD (polychlorinated dibenzo-*p*-dioxin), PCDF (polychlorinated dibenzofuran) and PCB (polychlorinated biphenyl) is assigned a single scaling factor termed the TEF. By definition, the TEF for TCDD (the index chemical for DLCs) is 1.0. To apply TEFs to an environmental mixture of DLCs, each individual compound's exposure concentration is multiplied by its specific TEF, yielding the individual PCDD, PCDF or PCB dose that is equivalent to a dose of the index chemical. These index chemical equivalent doses are then summed. To estimate risk associated with the mixture, the dose-response function for the index chemical is evaluated at this sum, which is an estimate of the total index chemical equivalent dose for the mixture components being considered.

Within the EU, the chemical regulatory regime that is most similar to that for pesticides is the one for biocides. Under the biocides' scheme, the general health-based reference value is the systemic *Acceptable Exposure Level* (*AEL*). The term AEL resembles the AOEL, but the omission of the term operator emphasises that the AEL is the reference value for the human population as a whole. The starting point for the AEL is a NOAEL / LOAEL / BMDL, to which assessment factors are then applied. Since AELs are systemic reference values, they should in principle be derived independently of the route of exposure. Such AELs represent the internal (absorbed) dose available for systemic distribution from any route of exposure and are expressed as internal levels (mg/kg bw/day). The Technical Notes for Guidance for Quantitative Risk Characterisation[22] (currently being updated to

[21] http://www.epa.gov/risk_assessment/guidance.htm
[22]http://ihcp.jrc.ec.europa.eu/our_activities/public-health/risk_assessment_of_Biocides/doc/TNsG/TNsG_ANNEX_I_INCLUSION/Revision_TNsG_Annex_I_Inclusion_Chapter_4.1_2009.pdf

reflect the introduction of the Biocidal Products Regulation 528/2012) propose that three AELs be derived: acute, medium-term and long-term. For the acute AEL, the estimated duration of human exposure is ≤ 24 hours, and the NOAEL is selected from a single-dose study or a repeated dose study that demonstrates relevant acute effects. For the medium-term AEL, the estimated duration of human exposure is 24 hours to 3 months (maximum 6 months) and the NOAEL is identified from a suitable repeated-dose study. The long-term AEL usually relates to human exposure of greater than 6 months (minimum three months) and is derived from a NOAEL identified in a chronic or repeated-dose study that demonstrates relevant chronic effects. Assessment factors are then applied to the identified point of departure (NOAEL / LOAEL / BMDL). In addition to the standard inter-and intra-species differences, additional factors may be applied based on the nature and severity of the effect; the human (sub-)population exposed; duration extrapolation (for example, from sub-chronic to chronic); dose-response relationship (e.g., extrapolation from LOAEL to NOAEL, the slope of the dose-response relationship); and the overall quality of the data package.

Another major, and relatively recent, piece of chemicals' legislation in the EU is REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals, Regulation 1907/2006). REACH has introduced the concept of *DNELs* (*derived no-effect levels*) and *DMELs* (*derived minimal-effect levels*) to the human health risk assessment of 'industrial' chemicals (ECHA, 2012). DNELs provide exposure standards for threshold effects to be used in risk characterisation and hazard communication (not legally-binding exposure limits), whereas DMELs are a quantitative reference level to aid in the risk assessment of non-threshold effects. Depending on the available information and the exposure scenarios[23], it might be necessary to identify different DNELs for each relevant population and for different routes of exposure, as follows: for workers (inhalation and dermal); for consumers (oral, inhalation and/or dermal, depending on exposure patterns); for the general population (oral); and for susceptible groups (e.g. pregnant women, if appropriate). The point of departure for the DNEL is the NOAEL / LOAEL / BMDL. This is then modified, if necessary, to take account of route-to-route extrapolation (for systemic effects) and/or exposure duration, followed by the application of assessment factors in a way similar to those for the derivation of biocide AELs (but with an intra-species factor of 5 for workers compared with 10 for the general population). An interesting aspect in the derivation of DNELs is that allometric scaling is applied for inter-species extrapolation (see task 6). A DNEL is derived for each end-point, and the lowest DNEL is then selected. There are two important points to note in relation to DNELs: they are a long-term value, since it is considered that a DNEL that is protective for long-term exposure will also be protective for short- and medium-term exposures (although, in exceptional cases, an acute 15-minute DNEL can be derived only for the inhalation route where there are peak exposures); and, they are external values that relate to, for example, how much chemical there is in the atmosphere that can be inhaled or how much is present on the skin, not to how much is absorbed systemically.

There are two approaches to the derivation of a DMEL: 1), the 'large assessment factor' approach, as developed by the Scientific Committee of EFSA (EFSA, 2005; see also task 6, margin of exposure); and 2), the 'linearised' approach. In the former, a large assessment factor (10,000) is applied to a (modified) toxicological reference point (e.g., a BMD10 or BMDL10). In the latter, linear extrapolation is applied from the T25 to a dose/exposure that is associated with an acceptable lifetime excess risk (e.g., 10^{-5} or 10^{-6}). Different regulatory authorities have varying preferences for these methods.

Indicative Occupational Exposure Limit Values (*IOELVs*) are health-based limits set under the Chemical Agents Directive (98/24/EC). The Scientific Committee on Occupational Exposure Limits

[23] Within REACH, an exposure scenario is a set of conditions, including operational conditions and risk management measures that describe how the substance is manufactured or used during its life-cycle and how exposure of humans and the environment is controlled.

(SCOEL) advises the European Commission on the limits. The Directives[24] listing these limits require EU member states to introduce domestic limits for these chemicals, which must take account of the IOELVs. *Workplace Exposure Limits* (*WELs*) are legally binding British occupational exposure limits and are set in order to help protect the health of workers (Health and Safety Executive, 2011). WELs are concentrations of hazardous substances in the air, averaged over a specified period of time, referred to as a time-weighted average (TWA). Two time periods are used: long-term (8 hours); and short-term (15 minutes). The long-term (8-hour TWA) exposure limit is intended to control such effects by restricting the total intake by inhalation over one or more work shifts, depending on the length of the shift. Short-term exposure limits (STEL - usually 15 minutes) may be applied to control effects seen after brief exposures, such as eye irritation.

In Germany, *MAK values* (*maximum workplace concentrations*) and *BAT values* (*biological tolerance value for occupational exposure*) are established for the classification of carcinogenic, embryotoxic/fetotoxic substances and germ cell mutagens, and for the evaluation of measurement methods. The MAK value is defined as the maximum concentration of a chemical substance in the workplace air which generally does not have known adverse effects on the health of employees nor causes unreasonable annoyance (e.g., by nauseous odour) and are also 8-hour TWA values. Known effects of a substance in man are given highest priority in the derivation of the MAK value, which is based on the NOAEL for the most sensitive effect with relevance to health. If a NOAEL cannot be derived from the available data, a MAK value is not established.

Information on the occupational exposure limit systems in other EU member states can be accessed from the website of the European Agency for Safety and Health at Work[25].

The US National Advisory Committee for the Development of Acute Exposure Guideline Levels for Hazardous Substances (AEGL Committee) is involved in developing guidelines for the setting of *Acute Exposure Guideline Levels* (*AEGLs*)[26]. AEGLs are intended to describe the risk to humans resulting from once-in-a-lifetime, or rare, short-term exposure to airborne chemicals. They have been developed for use in emergency planning and prevention as well as during real-time emergency response actions in relation to the manufacture, processing, storage and transportation of chemicals and cleaning-up pollution. The context is both accidental and terrorist releases of chemical substances. Three AEGL values are defined for a substance in terms of the airborne concentration at which the general population would experience the following: AEGL 1: notable discomfort; AEGL 2: irreversible effects or have impaired ability to escape; AEGL 3: life-threatening effects or death.

An EU-funded research project has investigated the potential of *Acute Exposure Threshold Levels* (AETLs) to complement the AEGL approach whilst meeting needs specific to European users within the context of EU Seveso II Directive (COMAH). AETLs will define the exposure conditions in terms of airborne concentration and exposure time that will produce a series of specified levels of harm to people for a number of toxic chemicals[27,28]. Their anticipated use is to aid decision making within EU Member States on emergency planning and land-use planning as appropriate in relation to Seveso II sites (not risks associated with transport, the military or terrorism). In the AETL approach, hazard is used as a simple surrogate for risk. 'Named carcinogens' (which are thought to pose a risk of carcinogenicity after single exposure and are thus acutely toxic) will be ranked according to export judgement. For substances that are Toxic, Very Toxic, Irritant or Corrosive, the hazard measures are based on the substances' physicochemical and toxicological hazardous properties (estimate of the $4hLC_{50}$ as an indication of relative toxicity) together with tonnage. Currently, different EU member

[24] Directives 91/322/EEC, 2000/39/EC, 2006/15/EC and 2009/161/EU.
[25] https://osha.europa.eu/en/topics/ds/oel/members.stm
[26] http://www.epa.gov/oppt/aegl/
[27] http://www.ineris.fr/centredoc/TGD_06DR055.pdf
[28] http://mahb.jrc.it/index.php?id=45

states use different means to assess and express major hazards. For example, in Great Britain, the Health and Safety Executive (HSE) makes assessments of the *Dangerous Toxic Load* (*DTL*) for *Specified Level of Toxicity* (*SLOT*) and *Significant Likelihood of Death* (*SLOD*)[29]. The DTL describes the exposure conditions, in terms of airborne concentration and duration of exposure, which would produce a particular level of toxicity in the general population. The criteria that define a SLOT are fairly broad in scope and, importantly, are also relatively easy for non-scientists to understand in terms of the overall health impact. these two factors are used to calculate the Toxic Load. As for AETLs, animal 4hLC_{50} data is used as an indicator of relative toxicity.

Directive 2008/50/EC on ambient air quality and cleaner air for Europe sets legally-binding limit values for concentrations in outdoor air of major air pollutants that impact public health, such as particulate matter (PM_{10} and $PM_{2.5}$) and nitrogen dioxide (NO_2). It also sets target values, which are to be attained where possible by taking all necessary measures not entailing disproportionate costs. Air Quality Standards are concentrations recorded over a given time period that are considered to be acceptable in terms of what is scientifically known about the effects of each pollutant on health and on the environment. They can also be used as a benchmark to indicate whether air pollution is getting better or worse. An exceedence is a period of time (defined for each standard) where the concentration is higher than that set out in the Standard. Limit values are set for individual pollutants and comprise a concentration value, an averaging time over which it is to be measured, the number of exceedences allowed per year, if any, and a date by which it must be achieved. Some pollutants have more than one limit value covering different endpoints or averaging times.

Rather than risk management, pharmaceuticals undergo benefit-risk management. For medicinal products, the *therapeutic index* (*TI*) is the ratio between effectiveness of a dose and safety. In humans, the TI is calculated as the toxic dose (TD) for 50% of the patients tested (determined by clinical trial studies) divided by the effective dose (ED) that works for 50% of the patients (TD50 / ED50). If the TI is calculated from animal data, the ratio is LD50 / ED50. A higher TI indicates a greater margin of safety, since the dose needed to be effective is much smaller than the dose that causes toxicity.

Some initiatives have a scope that extends beyond a single regulatory regime and global region. These do not involve the setting of reference values, but can be used in the evaluation of data on chemicals. For example, groups such as RISK21 (http://www.hesiglobal.org/i4a/pages/Index.cfm?pageID=3546) are looking to use new approaches to improve chemical risk assessments. Part of the approach proposed is to use more targeted mechanism-based data generation rather than the current approach of performing a set battery of studies irrespective of the toxicity profile of the compound. This is a development of the scheme outlined by the National Research Council of the National Academy of Sciences' vision of toxicology in the 21st century (NRC, 2007). Both these initiatives aim to take toxicity testing forward and improve the understanding of how environmental chemicals can affect human health, by developing and using new techniques and approaches. The intention is to have a range of computational and in vitro techniques based on human biology to identify potential effects on key toxicity pathways. This should result in a more focussed an mechanistic approach to toxicology and moves away from animal testing to *in silico* and *in vitro* methodologies combined with human population biomonitoring. These techniques are still being developed and are not yet validated for use as regulatory alternatives to the derivation of formal health based guidance values.

Two further programmes that are not specific to a particular regulatory regime have taken a structured approach to the assessment of data on chemicals and how this informs on their mode(s) of action. The WHO IPCS (International Programme on Chemical Safety) has developed a conceptual framework for the evaluation of the relevance of a cancer (Boobis *et al*., 2006) and a non-cancer mode of action for humans (Boobis *et al*., 2008). The pesticide thiazopyr, which induces rat thyroid follicular cell

[29] http://www.hse.gov.uk/chemicals/haztox.htm

tumours, was used as a case study in the context of the cancer mode of action framework (see task 5; Dellarco *et al*, 2006). The OECD has issued guidance on the development of adverse outcome pathways (AOPs; OECD, 2012a). AOPs provide a means to portray the existing knowledge concerning the pathway linkage between a molecular initiating event and an adverse outcome that is relevant to a regulatory decision. An example use of the AOP has been published for skin sensitisation, which has a defined molecular initiating event (covalent binding to proteins) (OECD, 2012b,c). Whilst such conceptual frameworks and pathways can be helpful in formulating a proposed mode of action, they are likely to be of limited use in those evaluations for which only the standard data set is available, since these would not routinely include mechanistic studies. Additionally, there would have to be evidence that the mode of action was not relevant to humans for the effects to be dismissed in the setting of a point of departure for guidance value derivation. Two specific examples of when this might be possible (thyroid effects, particularly tumours, in rats; and renal effects in male rats) are expanded upon in task 5.

Recently, the EPA has issued a draft conceptual framework for human health risk assessment to inform decision making by the agency (EPA, 2012b). The framework includes issues to consider, provides suggested questions to ask during risk assessment planning and execution, and identifies some useful practices.

4.2. Results of Questionnaire

A questionnaire was sent to regulatory authorities around the world, seeking information on the types of human health risk assessments they performed for chemicals and what approaches they adopted to deriving reference doses. The response rate has been poor with only 11 agencies responding. The majority of agencies are from Europe, which will skew the pattern of responses in the database as these agencies will be operating under similar legislative constraints and generally utilising the same guidance documents. A summary table of responses is in the supporting spreadsheet 'combined responses' described at Annex 2. Relevant guidance documents were obtained and evaluated.

The overall conclusions are:

- the predominant approach for reference dose setting is the use of a NOAEL and a 100 fold safety factor;
- dietary reference doses were set for acute and chronic (lifetime) exposures; non-dietary reference doses were typically for mid-term duration, although some agencies set non-dietary reference doses for acute, mid-term and long-term consistent with the EU biocides' approach;
- for continuous data NOAELs are determined based on a mixture of defined levels of change (e.g. 10%) and statistical significance;
- there is no general approach to clinical chemistry findings;
- the relevance to humans of findings such as liver hypertrophy is often dependent on a pattern of findings, not a single result;
- the criteria triggering the application of extra factors are reasonably consistent. However the value of any extra factor is determined case-by-case;
- chemical specific assessment factors (CSAF) had only limited use and appear to require significant levels of supporting data;
- the food contaminants' area was more willing to adopt alternative approaches than the biocide / pesticide area.
- BMD or BMDL is not used for pesticides or biocides in the EU. In North America it is used in specific cases, not routinely;
- BMD response levels (BMR) vary with the type of end-point being considered, but a 95% confidence level was common;
- the USEPA BMD software was the most commonly used;
- allometric scaling was not used routinely; one agency reported this option had been removed by legislative changes. The USEPA is looking to use allometric scaling in the future;
- route (non-oral) specific tests were not widely available;
- some agencies would set duration- / task-specific values but others just had a single non-dietary reference dose;
- uncertainty was frequently not mentioned in evaluations unless there was a particular reason to do so. When uncertainty was considered, it was usually as a simple text statement. There was no use of a fully quantitative approach to uncertainty such as the production of a mean value confidence interval.

For both dietary and non-dietary reference doses the general approach is to use an oral study NOAEL and apply a safety factor (100-fold default). Alternative approaches such as the BMD, allometric scaling or PBPK are used by some agencies in special circumstances.

5. EFSA TASK 4:

Collect and evaluate possible alternative reference points (to NOAEL and LOAEL) to be used for the derivation of AOELs, AAOELs and ADIs with a particular emphasis on the evaluation of the BMD approach and duly taking into account relevant EFSA publications.

Literature searches were performed by information scientists using the search criteria described in Annex 5. Based on the abstracts obtained, potentially relevant references were identified by the toxicology specialists, obtained and evaluated. The toxicologists also performed targeted online searches and obtained copies of references cited in core texts and primary references. In addition relevant responses to the questionnaire (see task 3) were taken into account.

Current and alternative reference points identified in the public literature have been considered in detail below.

5.1. Current reference points – NOAELs and LOAELs

It is generally agreed that many of the adverse health effects caused by substances are not expressed until the chemical, or an active metabolite, reaches a threshold concentration in the relevant organ. Whether or not this threshold concentration is reached is related to the level of exposure of the organism (human or test animal) to the substance: for a given route of exposure, there will be a threshold exposure level which must be attained before effects are induced. The threshold exposure dose or concentration may vary considerably for different routes of exposure, and for different species because of differences in toxicokinetics and possibly also in mechanisms of action. The observed threshold dose in a toxicity test will be influenced by the sensitivity of the test system and is a surrogate for the true threshold.

The No Observed Adverse Effect level (NOAEL) identified in a particular test will be simply the highest dose level or concentration of the substance used in that test at which no statistically (or toxicologically) significant adverse effects were observed (as judged by a statistical test and expert judgement), i.e. it is an operational value derived from a limited test. For example, if the dose levels of 200, 50, 10 and 5 mg/kg/day of a substance were used in a test and adverse effects were observed at 200 and 50 mg/kg/day but not at 10 or 5 mg/kg/day, the derived NOAEL would be 10 mg/kg/day. Thus, the NOAEL and LOAEL (lowest observed adverse effect level) values for a given study will depend on the experimental study design, i.e. the selection of dose levels and the spacing between doses. In the above example, the true NOAEL (determined if a large number of dose levels with small incremental changes had been used), might be 48 mg/kg/day or 12 mg/kg/day, but in either case the value from the study would be 10 mg/kg/day. The derivation of the NOAEL is dependent on the power of the study to 'observe' a significant adverse effect. For example, in a chronic rat study with a group size of 50, an increase in incidence of a lesion from 0 in controls to 6 in a test group (0 to 12%) would be statistically significant ($p=0.027$, Fisher exact test, two tailed); for a 90 day rat study (10 per group), an increase from 0 to 5 (0 to 50%) would be statistically significant ($p=0.033$, Fisher exact test, two tailed); for a typical dog study with a group size of 4, an increase from 0 to 3 (0 to 75%) would not be statistically significant ($p=0.14$ Fisher exact test, two tailed). Because of the low statistical power, for many dog studies expert judgement is used rather than statistical significance to determine NOAELs.

If there are several studies addressing the same effects from which different NOAELs could be derived, normally the lowest relevant value should be used in deriving the health-based reference

value. When it is not possible to identify the NOAEL, the "lowest observed adverse effect level" (LOAEL) is generally used as the point of departure.

The sensitivity of a study (which is related to the toxicological endpoint, the potency of the toxic substance, the exposure period and frequency, the variability within the species, the number of dose groups and the number of animals per dose group) may limit the extent to which it is possible to derive a reliable NOAEL from a particular test.

Although health-based guidance values are often set on the basis of a NOAEL, the NOAEL is not identical to the threshold of toxicity. There needs to be a certain amount of response before a difference between the responses of the control group and the test group can be identified. Modelling of the possible magnitude of response in toxicity studies has indicated that there can be about a 5% response at the NOAEL (Gaylor 1992). Dose spacing adds further uncertainty as there is uncertainty regarding where the NOAEL lies on the curve in relation to the actual threshold of response. The dose taken as the NOAEL (i.e. the highest tested dose level below the lowest tested dose to have an adverse effect) will be lower than the highest dose that would not cause an observable adverse effect if all possible doses had been tested. In addition, there is uncertainty about the significance of exposures less or greater than, but close to, the NOAEL.

Uncertainty is increased because of the differing practices of toxicologists in defining the NOAEL. The NOAEL is frequently considered to be the highest dose where no adverse effect occurs as defined by a pairwise statistical test between the test group(s) and the controls. There may be disagreement in how to proceed if the dose-group below that which shows a statistically significant pairwise difference from the controls shows a non-statistically significant difference. A p-value of 0.05 is usually regarded as the cut-off point for statistical significance (ie. $p<0.05$ is statistically significant).

The NOAEL depends critically on study design, the sensitivity of measurements of toxic endpoints, choice of doses, dose spacing and group size (statistical power). Thus two studies on the same chemical that are identical in every respect except the doses used can identify different NOAELs, because dose-spacing is a major determinant of the NOAEL. The existence of a dose-response relationship increases the confidence in the NOAEL, although dose-response relationships are not fundamental for defining a NOAEL (in contrast to the benchmark dose level, for which a dose-response is critical).

The NOAEL is itself the subject of statistical uncertainty and its reliability depends upon the power of the study. Confidence in the NOAEL could be increased by use of larger group sizes or smaller dose spacing (e.g. by the use of more groups at different dose levels).

Experiments that use fewer animals tend to result in higher NOAELs associated with greater imprecision in the determined NOAEL (i.e. increase the chance of a false negative [type II error] at any particular dose level). Thus, a NOAEL derived using a small number of animals per dose group results in additional uncertainty about whether the NOAEL is actually below the true threshold of response (see Brown and Erdreich, 1989). However, the selection of animal numbers per dose group is influenced by external drivers other than scientific ones regarding power in defining the NOAEL. Notably, the use of 4 or 6 dogs of each sex per dose group (on animal welfare grounds) makes rigid adherence to formal statistical analysis inappropriate. Often, it is necessary to consider either trends, or the responses in individual animals, against the spectrum of effects observed in the study in question and other studies before drawing inferences. A further problem is that the multiple endpoints studied in a typical toxicological study increase the possibility of type I errors (false positives). Yet another question raised by the concept of the NOAEL is what constitutes an adverse effect.

The LOAEL is subject to many of the statistical problems discussed above in relation to the NOAEL. As with the NOAEL, using a LOAEL means the information about the shape of the dose-response curve is not used, although the steepness of the curve should be taken into account when selecting an additional uncertainty factor to account for using a LOAEL rather than a NOAEL. Thus, when extrapolating from a LOAEL to a NOAEL there is uncertainty about how close the resulting figure is to the actual threshold for an effect. As is discussed below, where a health-based guidance value, such as an ADI or TDI, is defined by a LOAEL, an extra uncertainty factor of up to 10, most frequently 3, is often used.

5.2. Alternative reference points

5.2.1. Benchmark Dose

This has been reviewed extensively in a recent EFSA Scientific Opinion (EFSA, 2009b)

It is recognised that the NOAEL is not very accurate with respect to the degree to which it corresponds with the (unknown) true threshold of adversity/toxicity. It is important to understand that the NOAEL is the dose at which the investigator is unable to demonstrate an effect; this is not the same as proving that there really was no effect at that dose. Also, only the data obtained at one dose level (NOAEL) are used (in a quantitative manner) rather than the complete dose-response data set.

In response to the general call for consideration of the dose-response curve as a whole rather than to use only the data obtained at one dose level (NOAEL), alternatives for dose-response assessment have been proposed such as the benchmark dose (BMD) concept (Crump, 1984; Gaylor, 1988; USEPA, 1995; Slob and Pieters, 1998). The BMD methodology involves fitting a mathematical curve (equation) to the experimental dose-response data points and using all the plausible equations that fit the dataset to select a BMD. The BMD is the dose that results in a predetermined level (e.g. 5% or 10%) of adverse response, i.e. the critical effect size or benchmark response (BMR). The lower 95% confidence limit (BMDL) of the BMD is often taken as the starting point (‘point of departure’) for determining reference values. The presentation of the BMDL and BMDU and their ratio provides a measure of the uncertainty of the estimate.

It seemed that it was simply a matter of time before the BMD would replace the NOAEL as the regulatory tool of choice. Yet, more than 25 years since the initial development of BMD, the NOAEL is still the predominant approach in routine use. The reality is that all techniques have advantages and disadvantages, and a new approach will be adopted if it offers a favourable balance of these to the user community. In practice, the theoretical advantages of the BMD approach are often outweighed by the practical disadvantages it poses in a regulatory context (Travis et al., 2005). Some of these difficulties are highlighted in the practical aspects section below.

The main advantages of the BMD approach over the NOAEL are:

- the BMD makes extended quantitative use of the dose-response data from studies in experimental animals or from observational epidemiological studies (EFSA Scientific Opinion, 2009b), rather than utilising a single dose (i.e. the NOAEL or LOAEL);
- the BMD is independent of predefined dose levels and spacing of dose levels, resulting in a more consistent point of departure which reflects more accurately the true potency of the substance (as a consequence of the specified benchmark response) (EFSA Scientific Opinion, 2009b);

- the BMD approach provides a quantification of the uncertainties in the dose-response data (EFSA Scientific Opinion, 2009b);
- the BMD approach unveils the uncertainties in the response level hidden in the NOAEL, which is not a dose level with no effects (Slob, 1998);
- use of the BMD approach leads to a more precise and more transparent risk estimate (Slob, 1998), which, in turn, may lead to improved risk communication between risk assessors, risk managers, policy-makers and the public;
- the BMD takes into account the spread of the data at each dose level rather than relying on a mean;
- outlying values can be identified and excluded from the analysis.

In view of these strengths, EFSA Scientific Committee (2009b) concluded that the BMD approach is a scientifically more advanced method over the NOAEL method, and recommended that it becomes the preferred method in situations (i) where the identification of a NOAEL is uncertain (e.g. where only a LOAEL has been identified, but this is still below the BMD); (ii) when establishing the point of departure for substances that are both genotoxic and carcinogenic; and (iii) when performing dose-response assessment of observational epidemiological data. The Committee also recommended that the default values for uncertainty factors currently applied to the NOAEL remain appropriate when the point of departure is the BMD or the BMDL and that there is no need for any additional uncertainty factor because analyses show that reference values derived using the BMD or BMDL are very similar to those derived using the NOAEL (see also the case studies presented later in this report).

A perceived weakness of the BMD method relates to the reliability of the approach when results are obtained from toxicity studies performed according to the requirements defined in current testing guidelines. For the derivation of reliable dose-response relationships, the classical study design of three dose groups and a vehicle control group is not ideal, especially if one considers the unfavourable possibility that in a particular experiment, adverse effects may be identified only at the highest dose level. An improved benchmark model fit would be possible by increasing the number of dose groups without changing the total number of animals in the test. However, such a change in study design would generally no longer allow a proper derivation of a NOAEL. It is important to clarify that the current standard testing designs are not a major obstacle to the application of the BMD methodology as, although it is true that the BMD obtained from three dose groups rather than six dose groups is more uncertain (Wout Slob pers comm., 9/1/2013), the uncertainty will be reflected in the confidence limits (ratio of BMDL to BMDU). Furthermore, the same uncertainty applies to the NOAEL identified from any such study but the degree of uncertainty is hidden. Therefore, it is a misconception to conclude that certain datasets might not be suitable for BMD modelling. If a dataset is not suitable for the derivation of a BMD, then it is also not suitable for the derivation of a NOAEL or LOAEL or other point of departure (PoD).

Yet, there is a large number of practical obstacles to the increased use of the BMD methodology in a regulatory context, as follows.

- A huge range of dose-response models is available. Simplistically, as long as the model fits the data well, it should not much matter which model is chosen. But, if the BMD or BMDL is below the lowest dose (as could apply in scenario (i) in the EFSA opinion mentioned above) or higher than the highest dose, then the choice of the model will be critical, and the result may well depend more on the choice of model than on the data. One solution is to combine BMD estimates from different models, weighted according to the quality of fit of each model to the data (Travis et al., 2005). However, this can be resource-intensive unless built into the modelling software.

- There are no agreed critical effect sizes or critical benchmark responses for either continuous or quantal data for different effects and it is hard to imagine that international consensus could ever be arrived at in determining exactly what magnitude of each effect is to be considered adverse. The EFSA Scientific Panel (EFSA, 2009b) proposed 10% for quantal data and 5% for continuous data but this has not been agreed by EU risk managers. The USEPA uses 10% as a default but the value can vary depending on the end-point used. There is also variation in whether to use relative risk or extra risk.
- Currently there is a lack of statistical and modelling expertise among most risk assessment practitioners, but this is essential for a reliable interpretation of model outputs.
- The BMD analysis of non-standard study designs, unusual experiments and insufficiently reported publications may pose great difficulty. Current BMD software requires the input of the arithmetic mean and standard deviation for each treatment group when evaluating continuous data, and may also require the number of animals in each group. Ideally, data from individual animals should be used. Where the study design contains factors other than treatment, e.g. replicate investigations over time, these inputs may not be appropriate summaries of the data. In addition, missing values are a common feature of toxicity studies. The effect of missing values, particularly in small studies (e.g. studies in dogs) can be significant. For quantal data such as found in evaluations of developmental studies where nested approaches are used, there is a need for information on individual litters and pups, which is normally only available in study reports or raw data. The latter is not necessarily an issue when preparing a summary document from study reports, but prevents the use of nested models for analyses based on summary texts such as DARs.
- For each study, in order to identify the most sensitive effect driving the lowest BMD, the dose-response analysis should theoretically be performed for all the effects seen at or around the LOAEL as, depending on the slope of each effect, different BMD values may be obtained. This could be very time-consuming compared with the identification of the NOAEL. The approach of the USEPA is to perform BMD analyses on selected end-points and selected studies only.

Given all of these barriers, perhaps it is time to recognise that in the near future the BMD will not entirely replace the NOAEL in general use for the routine evaluation of existing studies. One practical way forward could be for NOAELs and expert judgement to guide an evaluator to the most critical study and critical endpoint for a given chemical, and at this point for the BMD method to be invoked as a higher tier or supplementary approach. For new studies, a BMD analysis could be included within the data / report generation software.

Despite all of these potential practical problems, the scientific supremacy of the BMD approach compared with the NOAEL method should be an incentive to apply it at least as a higher-tier or supplementary method when the critical study for the derivation of a reference value has been identified. The application of the BMD approach at this stage of the risk assessment process will generate a more robust and more transparent risk estimate with an indication of the associated uncertainty, which, in turn, could lead to improved risk communication between risk assessors, risk managers, policy-makers and the public. Such a tiered strategy has already been applied by EFSA and JECFA on several occasions when deriving reference values for contaminants.

Practical experience of modelling data using BMD software

A particular aspect identified in the project specification was to consider the practical use of BMD modelling. To investigate some of the practicalities of the BMD approach and to gain practical experience of such methodology, a number of datasets underlying the reference values (ADI, AOEL and ARfD) of 8 pesticide active substances were analysed. These 8 substances were selected following the exercise of reviewing the reference doses and EFSA conclusions (see activity 1):

- Carbofuran – BMD used in derivation of ref dose
- Cyanamide – LOAEL used
- Etridazole – LOAEL used
- Fenoxycarb – LOAEL & multiple studies / end-points used
- Propanil – LOAEL for RBC effects used
- Sulcotrione – LOAEL used but 100 factor maintained
- Triazoxide – LOAEL and steep dose response identified
- Fluapyroxad – multiple end-points identified.

Details of the outcomes of the BMD evaluations are presented in a summary table in Annex 3. Full details of the model outputs (numerical and graphical) are presented in Annex 4

In developing these case studies, the time taken to perform the analyses was recorded, to provide a comparison with the time taken to perform an equivalent analysis based on NOAELs.

The datasets were analysed using PROAST Graphical User Interface (GUI) (from RIVM, NL) and USEPA BMDS software (version 2.2). These software packages can both be downloaded for free from the web[30]. The evaluations performed with the USEPA BMDS software gave very variable results (up to 2 orders of magnitude) for the same dataset, depending on the dose-response model fitted to the data. CRD was unsuccessful in contacting the software developers at the USEPA and to obtain clarification on such outputs. The BMDS software was also found to be unreliable in that it would generate values with a dataset on one occasion but not when the same dataset was run subsequently; there were also difficulties in obtaining the graphical output of the curve fitting. Therefore, not knowing their reliability, CRD has only included minimal information on the USEPA BMDS estimates in the summary table.

Following initial difficulties with some data sets run on the PROAST GUI, CRD sought and obtained the help of its developer (Prof. Slob at RIVM). This ensured that the PROAST GUI estimates obtained could be considered reliable. In general, Prof. Slob produced similar results to those obtained initially by CRD. In addition, for some data sets he was able to run more appropriate models that were not available via the GUI.

The analyses performed were limited to 8 substances. Therefore, given the rather small number of datasets, no meaningful comparisons between the estimates of the NOAEL, LOAEL, BMD and BMDL values can be made. This was outside the scope of the exercise.

The following practical aspects and important theoretical considerations identified from the exercise.

1) The time taken for the analysis of a single dataset (1 study with up to 5-6 endpoints/effects) ranged from 1.5 – 15 hours. This value will probably decrease as users become familiar with the software.

2) Appropriate data (e.g. SD) were often not in the summary reports (e.g. DARs), making it difficult to perform BMD analyses retrospectively without access to the original data in the study reports. BMD software packages use the geometric mean and geometric standard deviation for continuous variables when processing the data, but these parameters are not normally presented in study reports. PROAST converts the arithmetic mean and standard deviation to a geometric mean and geometric standard deviation. The adjustment is normally

[30] PROAST = http://www.rivm.nl/en/Library/Scientific/Models/PROAST
BMDS = http://www.epa.gov/ncea/bmds/index.html

small if the data spread is small. This is one reason why data on individual animals are preferred to summary data.

3) Some level of statistical or modelling expertise is required to understand the significance of the outputs and how the software works.

4) The software is updated frequently; this is not a major problem, but minor changes may result. The GUI for PROAST has undergone significant changes recently to improve functionality but still does not give access to all the options available in the core (or menu) version of the programme.

5) A user-friendly software, that is intuitive and reliable, is essential before BMD analyses can be used by general evaluators.

6) Current software requires hands-on training, improved instruction manuals and better help functions.

7) On a statistical basis (sample size of standard toxicity studies), a BMR of 5% for continuous variables and an extra risk of 10% for quantal variables are considered the lowest response levels that can be reliably detected in a toxicity study. These response levels have therefore been proposed as the critical BMRs in order to match closely the estimates of standard toxicity study NOAELs. However, if for some effects, a higher response level is considered adverse (e.g. 20% for acetylcholinesterase inhibition) or in some studies the lowest dose has too high a response, then higher BMR values should be estimated. On the other hand, if for some effects, a lower response level is considered adverse (e.g. rare malformations or tumours or presence of methaemoglobin) or in some studies the lowest dose has too low a response, then lower BMR values should be estimated.

8) The BMD is an uncertain deterministic value and it differs among models (especially when the data are poor). The uncertainty range (or confidence interval) around the BMD indicates which values the true BMD might have. The uncertainty is the distance (ratio) between the BMDU and BMDL (also called imprecision factor = BMDU / BMDL).

9) BMD modelling is not a tool to identify the best fitted curve for a data set, but a tool to investigate all the plausible curves that fit the dataset. In the BMDS software the results of all the plausible curves are presented and the user chooses the most appropriate one; the EFSA opinion (EFSA, 2009b) suggests the curve giving the lowest BMDL value that cannot be discounted as unreliable should be used. PROAST runs multiple possible models and, since different models may give difference confidence intervals (model uncertainty), it provides a combined, overall BMDL-BMDU range based on those models that give plausible dose-response curves. The better the data, the less difference in the BMD values and confidence intervals obtained from the different models (less model uncertainty) will be and less wide the overall confidence interval.

10) The BMDL is always a more robust PoD than the NOAEL.

11) If the data are poor and no BMDL can be derived (e.g. it is zero), then the NOAEL / LOAEL identified from that dataset is also poor/unreliable. Therefore, the dataset is not suitable for deriving a PoD. A large imprecision factor also indicates that the data are poor and not suitable for the identification of a PoD. It is a misconception that such datasets are not suitable for modelling. In both these cases, the correct interpretation is that the data are not suitable for the derivation of a robust PoD. Several reasons for large imprecision factors or a value of zero for a BMDL were observed in the case studies; these included a high background rate with a small increase in response; large within-group variation; 100% response at the lowest dose; the test values being different from the control values for reasons other than the dose (e.g. pre-dosing differences); and possible outliers influencing the statistical significance of the results. BMD modelling can thus help to identify datasets that

are inadequate for risk assessment. Conversely, it helps to identify good-quality datasets that are suitable for risk assessment and from which a reliable PoD can be identified.

12) BMD analysis identifies if co-variates (e.g., sex) are statistically significantly different from one another, and so should be analysed separately, or if they can be combined to reduce the uncertainty in the result.

To gain a potential insight into how a move from using a NOAEL to a BMD_L might impact on the need to review existing values a comparison was made for the case-study compounds between the BMD_L and its equivalent NOAEL / LOAEL. The results of this limited analysis are outlined in Table 6. The indications are that where reliable BMD_L values are obtained the variation between the NOAEL and BMD_L is within an order of magnitude and within typical dose spacing in regulatory toxicity studies (see tasks 1 & 2). Therefore the use of a BMD approach is not inconsistent with one based on NOAELs and would not automatically require a retrospective review of previous evaluations.

Table 6 - Summary table of BMDL (lowest reliable value from PROAST) and NOAELs

Active substance	**NOAEL (mg/kg bw/d)**	**Lowest BMDL (mg/kg bw/d)**	**Ratio (NOAEL : BMD_L)**
Carbofuran ADI & ARfD	<0.03	0.006	5
Carbofuran AOEL	0.03	0.015	2
Cyanamide ADI & AOEL	<0.06	0.03	2
Cyanamide ARfD	5	0.9	6.5
Etridiazole ADI	<5	0.0005	Not reliable
Etridiazole AOEL	3.1	1.09	3
Fenoxycarb ADI	<5.3	$<10^{-5}$	Not reliable
Fenoxycarb AOEL	9.7	5.4	1.8
Fluxapyroxad ADI	2.1	3.3	0.6
Fluxapyroxad AOEL	6.0	24	0.25
Fluxapyroxad ARfD	25	25.3	1
Propanil ADI & AOEL	<5	6.8	0.8
Propanil ARfD	7	6.3	1.1
Sulcotrione ADI	<0.04	No reliable value	
Sulcotrione AOEL	0.06	0.015	4
Triazoxide ADI	<0.05	2.6	52
Triazoxide AOEL	0.2	0.20	1

Conclusions on BMD

- BMD modelling provides a superior scientific approach to the NOAEL to derive a PoD for a reference dose, as it makes use of all the dose-response data and provides an indication of the uncertainty (confidence) hidden in the PoD.

- BMD modelling is not at present a tool for routine use by human health risk assessors. The software programmes are still being developed, can be unreliable and they require a degree of specialist knowledge for reliable interpretation.
- However, despite these practical problems, the scientific supremacy of the BMD approach compared with the NOAEL method should be an incentive to apply it at least as a higher-tier or supplementary method when the critical study for the derivation of a reference value has been identified. The application of the BMD approach at this stage of the risk assessment process will generate a more robust and more transparent risk estimate with an indication of the associated uncertainty, which, in turn, could lead to improved risk communication between risk assessors, risk managers, policy-makers and the public.
- BMD modelling can be used to supplement a NOAEL derived in the conventional way by providing confidence limits. It could also be used to give an indication of how large an additional factor might be appropriate to apply to a LOAEL.

5.2.2. Threshold of toxicological concern (TTC)

The threshold of toxicological concern (TTC) is a surrogate for a specific health based guidance for a chemical. A generic threshold of acceptable exposure is determined based on the chemical structure and properties of the molecule (Kroes et al, 2004). The TTC is of value for performing risk assessments for data poor compounds such as metabolites or degradation products found in the environment and for chemicals where exposures are relatively low. In such cases the generic values of the TTC can be used as the basis for a risk assessment, without requiring the generation of data from animal studies. The TTC requires a reliable estimate of potential human exposures so can have limitations where exposures are not well controlled or exposure data are limited. A detailed consideration of the TTC approach was performed for EFSA contract EFSA/PPR/2008/01 (http://www.efsa.europa.eu/en/supporting/pub/44e.htm). The use of the TTC has been favourably reviewed recently by the EFSA scientific committee (EFSA, 2012a) and the EFSA panel on plant protection products and their residues (EFSA, 2012b) .

5.2.3. ED10, TD50 and T25

Other possible toxicological starting points are: the effective dose 10 - ED10 (the dose producing a 10% increase in an adverse effect, related to the control response); the tumourigenic dose rate 50 - TD50 (the chronic dose rate which would halve the percentage of tumour-free animals at the end of the standard lifespan for that species); and the tumourigenic dose 25 – T25 (the chronic daily dose which produces a 25% of the animal's tumours at a specific site, after correction for the spontaneous incidences within the standard lifespan of that species). These are analogous to the BMD approach and are covered under the linear extrapolation and MoE approaches in Task 6.

6. EFSA TASK 5:

Collect and evaluate relevant scientific literature or other relevant publications or information sources (e.g. guidance documents, government reports) related to the understanding or assessment of the adverse or adaptive nature of effects observed in toxicological studies and provide criteria to evaluate the appropriateness of toxicological effects to be used for derivation of reference points for setting health-based guidance values.

Literature searches were performed by information scientists using the search criteria described in Annex 5. Based on the abstracts obtained, potentially relevant references were identified by the toxicology specialists, obtained and evaluated. The toxicologists also performed targeted online searches and obtained copies of references cited in core texts and primary references. In order to be of relevance to health based guidance value setting the analyses concentrated on end-points identified as critical from the evaluations performed in Tasks 1 & 2:

Body weight
Liver weight and hypertrophy
Kidney effects in male rats
Thyroid effects in rats
General organ weight changes
Acetylcholinesterase (AChE) inhibition
Clinical chemistry/haematology changes

6.1. Body Weight effects

Effects on body weight (usually reductions) are observed routinely in standard and non-standard toxicity studies conducted in experimental animals. These may be accompanied by reductions in food consumption and/or by other toxic effects.

It is often debated whether these effects should be considered genuine toxic effects of a substance and which magnitude of body weight reductions should be considered adverse.

Our analysis of the 224 pesticide active substances for which EFSA conclusions have been considered in this project, has shown that 32 (14%) were identified as having at least one reference value set exclusively on the basis of effects (at the LOAEL) on body weight.

Among these 32 substances, the reference value concerned was the ADI for 17 substances; the AOEL for 14 substances; and the ARfD for 13 substances. For some substances, two or even all three reference values were based on such effects.

For the 17 substances for which the ADI was based on such findings, reductions in body weight ranged from 4 to 24% of the control values. Reductions in body weight gain were higher, ranging from 15 to 52% of the control values. For the majority of the substances, such reductions were statistically significantly different from controls. Only 2 of the 17 substances considered (12%) had reductions in body weight < 10%. Reductions in body weights were associated with reductions in food consumption in approximately 2/3 of the substances considered. For 5 out of the 17 (29%) substances considered, no significant toxic effects other than reductions in body weight were observed up to the highest doses tested. In the remaining substances, additional toxic effects were noted at higher dose levels.

For the 14 substances for which the AOEL was based on such findings, reductions in body weight ranged from 6 to 18% of the control values. Reductions in body weight gain were higher, ranging from 15 to 85% of the control values. For the majority of the substances, such reductions were statistically significantly different from controls. Only one substance (out of the 14 - 7%) had reductions in body weight < 10%. Reductions in body weights were associated with reductions in food consumption in all except one of the substances considered. Only for two out of the 14 (14%) substances considered, no significant toxic effects other than reductions in body weight were observed up to the highest doses tested. These were studies conducted in the dog. In the remaining substances, additional toxic effects were noted at higher dose levels.

For the 13 substances for which the ARfD was based on such findings, 10 caused effects on maternal body weight in developmental toxicity studies in rabbits and/or rats. Reductions in body weight ranged from 1.4 to 8% of the control values. Reductions in body weight gain were higher, ranging from 17 to 86% of the control values. For 3 substances, there was severe body weight loss occurring in maternal animals during the first days of dosing (prenatal developmental toxicity studies). For the majority of the substances, such reductions/losses were statistically significantly different from controls. Reductions in body weights were associated with reductions in food consumption in all of the 13 substances considered. Only for two of the 13 (15%) substances considered, were no significant toxic effects other than reductions in body weight (in maternal animals of developmental toxicity studies) were observed up to the highest doses tested. However, even for these two substances, this pattern of effects is not considered unusual because investigations of maternal toxicity in prenatal developmental toxicity studies are very limited. In the remaining substances, additional toxic effects were noted at higher dose levels.

With regard to the toxicological significance of body weight effects, our analysis shows that in the overwhelming majority of cases, reductions in body weights were accompanied by reductions in food consumption. In those cases where food consumption was not affected, it is possible that the effect had not been reported because of lack of statistical significance or because food consumption of individual rodents in shared cages is difficult to determine accurately.

Also, in the overwhelming majority of cases, additional toxic effects were noted at higher dose levels. Therefore, the most obvious conclusion from this analysis is that the decrements in body weights observed were most likely due to the treated animals not “feeling well” and, hence, eating less. Thus, unless proven otherwise, reductions in body weights should be seen as a sensitive marker of toxicity. The only possible exception to this rule would be a situation where it has been clearly demonstrated that the decreased body weight effect in a dietary exposure study is secondary to palatability. If the effect were due to palatability, it is unlikely that humans would be similarly affected as the exposure would normally be >100 fold lower and thus less ‘repulsive’. Additionally, humans have a less sensitive olfactory system than most other mammals.

It has often been argued that effects on body weights observed as the secondary consequence of the local irritation of the substance in the gastro-intestinal tract (especially in dogs) should not be considered adverse. This is a rather poor argumentation in that the local nature/origin of the effect does not detract from its adversity.

With regard to which magnitude of body weight reductions should be considered adverse, our analysis shows that slightly different levels of response are considered adverse depending on whether the effect relates to body weight or body weight gain. For body weight effects, even a 1.4% decrement was considered adverse on one occasion, whilst for body weight gain, a 15% decrease was the lowest degree of response considered adverse. Overall, the results of the evaluation are consistent with a broad acceptance of a 10% reduction in body weight gain as the cut-off between adaptive and adverse

response. The situation is less clear with regard to an effect on body weight, where a specific cut-off value does not seem to exist. Here, a case-by-case approach should be taken, which accounts for other considerations, e.g. statistical significance, time-course of the effect, presence of other toxic effects, associated (and magnitude of) effect on body weight gain.

6.2. Liver effects

The liver is an organ that is always adapting to changes in the body of an animal. These adaptations include glycogen storage and release; responding to the need to metabolise xenobiotics; and synthesising proteins in response to blood loss or inflammation. These adaptive responses are beneficial in maintaining homeostasis but there is a point when the changes result in a functional impairment that affects the whole organism. The primary response of the liver when exposed to xenobiotics is to synthesise additional metabolic capacity, primarily in the form of cytochrome P450 enzymes (Cyp). The actual Cyp forms induced vary with the xenobiotic stressor, as does the underlying mechanism (Hinton et al, 2009). A result of this induction of metabolising enzymes is frequently, but not always (Hinton et al, 2009; Hall et al, 2012), an increase in the size / weight of the liver and a pathological change identified as 'hypertrophy'. Hypertrophy is defined as an enlargement of the liver cells / the accumulation of fluids – it is not an increase in cell number, which is defined as hyperplasia. Hypertrophy is a morphological description, not an indication of adversity *per se*.

As has been shown earlier in this report, in the analyses of the EFSA conclusions on pesticides, effects on the liver are critical in determining health-based reference values for many chemicals. A major discussion point is whether the findings in rodent livers are adaptive or adverse and whether are relevant or not to humans. The descriptions of hepatotoxicity in humans following chemical exposure generally include findings such as necrosis and fibrosis (Hinton et al, 2009). If these lesions were present in animal studies they would be considered as adverse, not adaptive. Many investigations of liver toxicity in animals concentrate on whether a mechanism of liver carcinogenicity is relevant to humans but there has been less work on determining the point at which an adaptive / reversible response should be considered as adverse for humans. For example, when the size of the liver is greatly increased there will be disruption of the blood flow and pressure on other organs (Hall et al, 2012). One of the complicating factors is the species differences in response to xenobiotics (Williams & Iatropoulos, 2002; Williams & Perrone, 1996).

Three guidance documents are available on the interpretation of liver hypertrophy or increased liver weight (USEPA, 2002; Andrew, 2005; WHO, 2006). These documents are broadly consistent in identifying findings that would not be considered adaptive i.e. necrosis, fibrosis, hyperplasia and marked increases in serum activities of markers of hepatic damage.

The need to consider the overall weight of evidence and not just a single finding is stressed in the documents. Andrew (2005) proposes that an increase in liver weight relative to body weight of less than 10% is not adverse in isolation; this approach is broadly consistent with recent decisions in PRAPeR meetings. None of the guidance documents give a value for increased liver weight that would be adverse in its own right. In a recent paper by Hall *et al.* (2012), it was considered that an increase in liver weight of 50% in short-term studies would be unlikely to be compatible with adequate survival in a chronic study. The increase in hepatic Cyp activity is not adverse in itself, but it can alter the metabolism of endogenous molecules or nutritive molecules. Therefore the WHO (2006) recommends that there should be no significant induction of xenobiotic-metabolising enzyme activity in experimental animals at doses below the health-based guidance values for humans. Alterations in clinical chemistry parameters such as serum activities of marker enzymes for hepatotoxicity are not necessarily associated with an adverse effect (Hall *et al*., 2012) and can be secondary to enzyme induction.

In addition to considering whether a finding of hepatotoxicity is adaptive or non-adverse in the animal model, there is also the issue of relevance of the mode / mechanism of action to humans. A number of mechanisms of action for hepatic hypertrophy have been identified, involving one or more nuclear receptors: pregnane X receptor (PXR); constitutive androstane receptor (CAR); peroxisome proliferator-activated receptor alpha (PPARα); aryl hydrocarbon receptor (AhR). It is generally accepted that, of these mechanisms of action, activation of the AhR results in a range of changes in the liver over and above hypertrophy (Maronpot *et al*., 2010; Hall et al, 2012) that should be considered adverse for humans. Hepatocellular tumours produced via PPARα have been considered to be of no relevance to humans due to the low responsiveness of the human liver to the proliferative response induced by this receptor (Andrew, 2005; ECHA, 2009; Hall et al, 2012).

It is proposed that a pragmatic scheme to provide a more consistent approach to the interpretation of liver hypertrophy and the derivation of a NOAEL for a particular study could be based on the following outline, taking account of existing guidance documents and accepted practice within PRAPeR:

- increases in relative liver weight of <10% are not adverse in isolation.
- changes in clinical chemistry parameters related to hepatic function / damage (e.g. marker enzymes such as alanine aminotransferase (ALT), γ-glutamyltranspeptidase (GGT), alkaline phosphatise (ALP); total bilirubin or cholesterol) within appropriate background control values are not adverse;
- isolated findings of hypertrophy or clinical chemistry changes outside these ranges are not adverse;
- induction of xenobiotic metabolising enzymes is not adverse per se but a NOAEL for enzyme induction should be above the ADI, AOEL or ARfD;
- all other histopathological findings, or patterns of findings, of hepatotoxicity or associated clinical chemistry changes associated with hepatotoxicity should be evaluated using a weight of evidence approach and the default assumption is that they are treated as adverse unless a mode / mechanism of action has been identified that demonstrates lack of human relevance. To assist in assessing human relevance the data should be presented in a structured way, e.g. in line with the scheme proposed by the the IPCS (International Programme on Chemical Safety; Boobis *et al.*, 2008).

6.3. Kidney effects in male rats

Chronic progressive nephropathy

Chronic progressive nephropathy (CPN) is a common, age-related renal disease that affects all conventional strains of rat used in pre-clinical toxicology studies, but in particular the most commonly used strains, Fischer 344 and Sprague-Dawley (Hard and Khan, 2004). It occurs in both sexes, but at higher incidences and with progressively greater severity in males than females.

The histopathological manifestations of CPN are basophilic tubules, thickened basement membranes, hyaline cast formation and glomerulosclerosis. Glomerular obsolescence is accompanied by mild interstitial fibrosis and focal accumulation of mononuclear inflammatory cells. Although it is

normally regarded as a disease of ageing rats, early lesions, in the form of tubular cytoplasmic basophilia and thickening basement membranes, have been reported in two-month-old male rats. It has therefore been suggested that CPN should be regarded not simply as a manifestation of the ageing process, but as a specific disease entity. Progression of the disease can ultimately lead to chronic renal failure. As well as being a degenerative disease, CPN is also a regenerative disease. Thus, histopathological findings consistent with regeneration may be observed, particularly in the later stages of the disease; these can include atypical tubule hyperplasia and very small, tightly coiled basophilic tubules arising within scattered atrophic areas, which have been referred to as 'small tubule regeneration' (Hard *et al.,* 1997). It has been suggested that certain chemicals can interact with CPN to increase the incidence of CPN-related proliferative lesions (Hard *et al.,* 1997; see below). Lesions of the renal arterioles and afferent glomerular vessels have not been observed as part of the CPN spectrum of changes; nor does it have an immunological basis. However, certain clinical chemistry changes in urine and serum do accompany CPN; these include proteinuria (specifically, albuminuria), which correlates with the degree of severity of CPN; hypoalbuminaemia; and a rise in serum cholesterol. Because CPN is a complex of a number of individual structural lesions, study pathologists are recommended to record the complex as a single entity, not as the individual elements of the disease process. However, if tubular basophilia, hyperplasia, casts or glomerular change occur in rats unrelated to the CPN disease state, such changes should be recorded as findings in their own right (Hard and Khan, 2004).

The aetiology of CPN is not known. However, a number of factors, primarily diet-related, have been identified as being able to influence the incidence and severity of the disease. Variation in the protein content of the diet has a particular effect: reduction in animal-derived protein is protective, whereas an increase in protein exacerbates the condition (Rao *et al*., 1993). Restriction of caloric intake is the most powerful dietary manipulation in reducing the disease process (refs to be added). The overall conclusion from the available information is that the total amount of food consumed by the rat during its lifetime is a determinant of the extent of spontaneous renal damage that occurs (Hard and Khan, 2004). This observation supports the hypothesis that CPN is a disease entity rather than a true ageing process. Because the high incidence of advanced CPN is a confounding factor in the interpretation of induced kidney changes in chronic toxicology studies, the NTP has (since 1994) adopted a lower-protein, higher fibre and fat diet for all its rodent studies; since the introduction of this diet, the severity of CPN has been reduced without major effects on growth or body weight.

There is no significant disease entity in humans that has the same features as CPN in the laboratory rat. In the rat, CPN progresses relentlessly, so that the incidence is virtually 100% by two years of age. Another feature is that it shows some direct dependence on dietary modification. In contrast, no specific kidney disease that is totally confined to the ageing kidney has been identified in humans. Additionally, the evidence is that a low-protein diet does not alter the progression of diseases that cause chronic renal failure in humans (Ruggenenti *et al.*, 2001). Moreover, the pattern of histopathological, inflammatory, vascular and clinical chemistry findings in human chronic renal diseases are not concordant with CPN. The prevailing view is therefore that CPN in rats has no strict human counterpart.

The UK pesticides' competent authority has recently considered how a NOAEL should be set for a two-year rat carcinogenicity study in which findings indicative of CPN were recorded. The lowest dose, with no renal effects, was 9 mg/kg/d. At 27 mg/kg/d, the renal findings were reported to be statistically significantly higher incidences of renal interstitial fibrosis and glomerulosclerosis in males. At the next dose of 83 mg/kg/d, the additional findings were further increased incidences of renal interstitial fibrosis and glomerulosclerosis in males, higher incidences of CPN in males, reduced female body weight gain and hepatic periportal fatty degeneration in males. The Notifier argued that the renal effects at 27 mg/kg/d were elements of CPN. It also argued thatgiven that the incidence of

CPN at this dose was within the laboratory historical control data, the kidney effects observed at 27 mg/kg/d were not treatment-related. Its bases for this conclusion were:

1. the interstitial fibrosis and glomerulosclerosis were recorded as individual disease elements separately from CPN (although the study pathologist regarded them as being related to CPN), but were associated to CPN as shown by the sums of the incidences of interstitial fibrosis and glomerulosclerosis and the incidences of CPN;
2. the mean severity grades of interstitial fibrosis and glomerulosclerosis were lower than those of the controls;
3. historical control data although highlighting considerable differences in the incidence of CPN, and by inference in the individual lesions, both between and within rat strains, showed that CPN at 27 mg/kg/d was a spontaneous lesion and;
4. the slighter higher food consumption by males throughout the duration of the study might have contributed to the generally slightly higher incidences of CPN and its associated lesions in the male treated groups.

These arguments were accepted by the UK CA and by experts from other EU Member States during the peer-review process and the NOAEL was set at 27 mg/kg/d. Overall, therefore, CPN and its associated lesions in male rats were dismissed as non-treatment related, spontaneous changes only at those dosing levels at which their incidences were within the laboratory historical control data.

Alpha$_2$-urinary globulin-associated nephropathy

Many chemicals tested in two-year rat carcinogenicity studies have been found to exacerbate the incidence and/or severity of CPN. In particular, chemicals that bind to alpha$_2$-urinary globulin (α_{2u}-globulin) are usually associated with a concomitant increase in the severity of CPN. It is now recognised that both aliphatic and aromatic compounds, representing a variety of solvents, fuels, pesticides, drugs and naturally occurring compunds can produce this toxicity (Swenberg and Lehman-McKeeman). Some of the chemicals that have undergone extensive mechanistic investigations in this regard include *d*-limonene, 2,4,4-trimethylpentane, unleaded gasoline, isophorone, 3,5,5-trimethylhexanoic acid and 1,4-dichlorobenzene.

Hepatic synthesis of alpha$_{2u}$-globulin occurs exclusively in male rats (not in female rats, or mice or humans of either sex). Synthesis of α2u-globulin is reported for female and male rats (but not other species) in the salivary, lachrymal, preputial, meibomian, and perianal glands (Mancini et al., 1989; Murty et al., 1987). The hormonal regulation of alpha$_{2u}$-globulin synthesis in each of these tissues is unique and, most importantly, not sex specific in the lachrymal, salivary, and preputial glands (Murty et al., 1987; MacInnes et al., 1986). The total amount of alpha$_{2u}$-globulin synthesis in female rats is <1% of that in males (Swenberg and Lehman-McKeeman, 1999). As a result, α_{2u}-globulin nephropathy is a renal syndrome that manifests only in male rats. This protein is a member of a superfamily of proteins that bind and transport small hydrophobic molecules. The rate-limiting step in the development of the syndrome is the reversible, but specific, binding of a chemical (or its metabolites) to α_{2u}-globulin. In male rat kidneys, α_{2u}-globulin is transferred from the plasma into the urine by glomerular filtration; it is then partially reabsorbed into the renal tubule cells where it is eventually broken down (Turkstra and van Raaij, 2001). It is hypothesised that the syndrome develops as a consequence of chemicals reversibly and non-covalently binding to the protein to form a complex that is more resistant to lysosomal degradation than the unreacted protein. This leads to protein overload, renal cell injury, compensatory cell proliferation and ultimately an increased incidence of renal tubule tumours (Swenberg and Lehman-McKeeman, 1999; Turkstra and van Raaij, 2001).

The induction of α_{2u}-globulin-associated nephropathy progresses through a specific, time-dependent sequence of pathological changes:

- *Within 24 hours of dosing*: rapid accumulation of hyaline droplets is observed in proximal tubule cells. The droplets contain α_{2u}-globulin (identified by immunohistochemical staining).
- *After 5 days of continuous chemical exposure*: single-cell necrosis and exfoliation in the P2 segment epithelium.
- *Following 3 to 6 weeks of continuous chemical exposure*: granular casts, formed from cellular debris, accumulate. Subsequently, tubule dilation occurs at the junction of the P3 segment and the thinner loop of Henle. Enhanced cell replication in response to cell death can be seen as increased cell division or as increased DNA synthesis, demonstrated by labelling techniques.
- *After prolonged chemical exposure (e.g., chronic toxicity studies)*: tubule hyperplasia, linear mineralisation in the renal papilla and renal tubular epithelial cell tumours are observed.

If treatment is stopped after the first three weeks of exposure, recovery occurs and normal renal structure is restored.

Overall, two conclusions can be drawn: 1) The sequence of events proposed to link α_{2u}-globulin accumulation to nephropathy and renal tubule tumours in male rats is scientifically plausible; and, 2) The α_{2u}-globulin-associated nephropathy appears to be unique to the male rat. It is now generally accepted, therefore, that the way in which the male rat kidney responds to chemicals that induce α_{2u}-globulin accumulation is probably not relevant to human risk assessment (Turkstra and van Raaij, 2001).

In order for the α_{2u}-globulin explanation for nephropathy and renal tumours to be accepted during a risk assessment, the USEPA (1991) and IARC (IARC, 1999) have defined certain criteria that must be met. The USEPA approach is that only some criteria must be fulfilled (2 and 3, below) and some additional information (1, 4-6, biochemical data, structure activity relationships) could be added. The IARC states that all the criteria 1-6 must be fulfilled.

Essential criteria α_{2u}-globulin MoA

1. *Non-genotoxic*. The agents and metabolites lack genotoxic activity based on an overall evaluation of *in vitro* and *in vivo* data.
2. *Induction of the characteristic sequence of histopathological changes in rat studies*. The sequence and timing of histopathological changes is outlined above. If a response is mild, not all of the lesions may be observed. However, some of these elements, including hyaline droplets, must be demonstrated to be present.
3. *Identification of the protein accumulating in tubule cells as α_{2u}-globulin*. Hyaline droplet accumulation is a non-specific response to protein overload in the renal tubule. It is therefore necessary to demonstrate the presence of α_{2u}-globulin within the droplets.
4. *Male rat specificity for nephropathy and renal tumours.* Clearly, since α_{2u}-globulin-associated nephropathy occurs exclusively in the male rat, this explanation cannot be the means to dismiss the relevance of findings to humans if they are also reported in female rats.

Additional information

5. *No nephropathy or renal tumours were induced in species other than the rat.* Positive responses in mice or other laboratory animals indicate that α_{2u}-globulin alone does not account for the nephropathy.

6. *Reversible binding of the chemical or metabolite to α_{2u}-globulin.* This can be shown in different ways, e.g. with *in vitro* and *in vivo* studies.

If, having given due consideration to the criteria above, it is concluded that a chemical has induced α_{2u}-globulin-associated nephropathy, the nephropathy are considered not to be toxicologically relevant. In such a study, the NOAEL is based on other, toxicologically relevant endpoints. The α_{2u}-globulin-associated nephropathy should not be used for setting a reference value, nor for human risk assessment (Turkstra and van Raaij, 2001). Severe nephropathy might eventually lead to changes in other parameters, such as body weight or urine volume; whether or not these are regarded as related to the nephropathy and thus not toxicologically relevant should be evaluated on a case-by-case basis.

Conclusion on α_{2u}-globulin

It is reasonable to dismiss certain renal effects in male rats as being of no relevance to humans, in which case they would not be used to set a NOAEL. However, the findings must fully meet the characteristics that define CPN or α_{2u}-globulin for these explanations to be accepted. In addition, for CPN, incidences must be within the laboratory historical control ranges. The evidence that nephropathy is explained by one of these modes of action should be fully explored and reasoned.

6.4. Thyroid effects in rats

Thyroid gland function is controlled by the hypothalamus and pituitary. The thyroid gland secretes the hormones thyroxine (T_4) and triiodothyronine (T_3) in response to stimulation by thyroid stimulating hormone (TSH), which is itself secreted by the anterior pituitary. TSH is regulated by both thyrotropin releasing hormone (TRH) secreted by the hypothalamus and by strong negative feedback in response to circulating levels of T_3 and T_4. Under normal physiological conditions, T_4 is secreted into the systemic circulation in greater quantities than T_3, such that circulating levels are approximately 10-times higher. T_3 and T_4 are transported around the body in strong (but not covalent) association with plasma protein. Humans, other primates and dogs possess a high-affinity binding protein for thyroid hormones, thyroxine binding globulin (TBG). TBG binds T_4 and, to a lesser extent, T_3. Rats, however, do not possess TBG. Both humans and rats possess low-affinity carrier proteins for thyroid hormones, thyroxine-binding prealbumin and albumin. Free and bound T_3 and T_4 are in dynamic equilibrium in the circulation, with more unbound T_3 than T_4.

A consequence of the lack of a high-affinity binding protein in rats is that more T_4 remains bound to lower affinity plasma proteins and so is more susceptible to removal from the blood, metabolism and excretion. As a result, the serum half-life of T_4 in rats is much shorter in rats (approximately 24 hours) than it is in humans (approximately 5-9 days). Likewise, the serum half-life of T_3 is also shorter in rats (approximately 6 hours) than in humans (approximately 24 hours). To compensate for this shorter half-life, the basal level of thyroid stimulation by TSH is much greater in rats than it is in humans. This difference between rats and humans in the basal level of stimulation manifests itself in the appearance of the follicles: in rats, they are relatively small and often surrounded by cuboidal epithelium; in humans, in contrast, they are less active and are large with abundant colloid, surrounded by relatively flattened epithelium. Overall, the rat thyroid gland is already chronically stimulated, and so slight perturbation in thyroid hormone levels from the administration of chemicals might lead to TSH levels above basal levels, which could readily move the follicles towards hypertrophy and hyperplasia. Consequently, rats are more sensitive to adverse effects on the thyroid than are humans. Mice and dogs appear to have an intermediate position with regards to species differences in thyroid function and potential disturbances: in some aspects they appear to be similar to rats, whereas in others they are more similar to humans. The most clearly defined species differences

therefore exist between humans and rats. In rats there are, moreover, intraspecies differences, since adult male rats have higher circulating TSH levels than females and are often more sensitive to goitrogenic stimulation and thyroid carcinogenesis. Additionally, follicular cells are often larger in male rats. In humans, there is no sex difference in TSH levels, but females develop thyroid cancer more frequently (Hill *et al*., 1998).

From the above it follows that TSH is the pivotal hormone responsible for chemically-induced excessive stimulation of the thyroid gland and is thus the central marker to be monitored in toxicology studies (van Raaij, 2001). The following three phases can be recorded when the stimulation of the thyroid by TSH is prolonged:

- *An initial phase (lasting several days)*: Rapid changes in thyroid morphology occur, including resorption of colloid from the follicular lumen, hypertrophy of follicular epithelial cells and an increase in vascularity.
- *Second phase of rapid growth:* A sustained increase in thyroid weight and size occurs. Follicular hypertrophy and hyperplasia can be detected.
- *Third phase of accumulation:* The growth of the thyroid reaches a plateau (there are limits to the extent to which the gland can increase in size and weight). Follicular hyperplasia may progress to nodular proliferation and eventually to neoplasia.

This progression is reversible, but the reversibility is dependent on the severity and especially the duration of the insult; if the process has progressed to the plateau stage, it cannot be reversed by the withdrawal of the causative agent. It should also be noted that these effects on the thyroid gland are the result of increased circulating levels of TSH rather than a direct effect of a chemical on the gland.

Since the regulation of TSH involves the hypothalamus, pituitary and the thyroid (HPT), disturbance of the HPT-axis can thus result in the pathological processes outlined above. There are a number of possible mechanisms by which non-genotoxic chemicals may induce thyroid hypertrophy and hyperplasia, and ultimately tumours, in rodents via a disturbance of the thyroid-pituitary axis.

1. *Inhibition of iodide uptake*

A number of anions are competitive inhibitors of iodide uptake. A particularly potent example is thiocyanate. These anions result in decreased circulating levels of T_3 and T_4 and a consequent increase in TSH production. The response of the thyroid to TSH is also increased. The relevance of this effect for humans is unclear and so it is regarded as possibly relevant.

2. *Inhibition of thyroid hormone synthesis*

Several classes of chemicals inhibit the process of thyroglobulin synthesis; thyroglobulin is a protein in the thyroid gland from which the thyroid hormones are synthesised. A mode of action in this category is inhibition of thyroid peroxidise (TPO), which catalyses the reaction of oxidised iodide ions with tyrosine residues at sites on the thyroglobulin molecule. The manifestations of this mode of action are the same as those for iodine deficiency. These effects are regarded as possibly relevant to humans.

3. Inhibition of 5'-monodeiodinase

The majority of total body T_3 is formed in peripheral tissues through the 5'-monodeiodination of T_4; inhibition of this enzyme can thus lead to an increase in TSH production to compensate for a decrease in serum T_3, and the negative feedback system that regulates TSH and responds to decreased serum (which acts through T_3 receptors) may be compromised. The characteristics of a chemical that acts through this route include increased serum levels of TSH and T_4 and decreased serum T_3. This mode

of action is relevant for humans, because, although excess T_4 is buffered by TBG, there may be effects on the negative feedback system in the pituitary that will cause thyroid stimulation.

4. Inhibition of thyroid hormone secretion

This seems to be an uncommon mode of action; lithium and excess iodide are well-known examples. The decreased serum T_3 and T_4 levels lead to an increase in TSH secretion. Since this is a direct effect on the thyroid, it is regarded as relevant to humans.

5. Liver-enzyme induction

Another mechanism whereby the thyroid-pituitary axis may be disturbed is liver enzyme induction; this can lead to increased conjugation and excretion of thyroid hormones, which in turn leads to increased thyroid stimulating hormone (TSH) secretion from the pituitary and then compensatory hyperplasia in the thyroid. Hepatocellular hypertrophy, as an indicator of an adaptive liver response, may also be observed. The induction of UDP-glucuronyl transferase, in particular, has been associated with thyroid tumours in rodents, since it is responsible for the metabolism of T4, which is compensated for by an increased production of TSH by the pituitary. In humans, the increase in metabolism would initially be compensated for by the reservoir of thyroid hormone bound to TBG. Strong inducers of P450 act by this mode. Where there is good evidence that thyroid toxicity is the result of this mode of action, it is generally regarded as not relevant to humans. However, there has been some speculation that the enhancement of thyroid hormone clearance via induction of conjugating enzymes in the liver hardly causes an increase in TSH levels (European Commission Specialised experts, 1999).

6. Modulation of TSH receptors

Substances that stimulate TSH receptors on the thyroid cause over-stimulation of the thyroid. As the cause of the stimulation is exogenous, the normal negative feedback mechanism is ineffective. Characteristically, serum TSH levels would be normal or below normal, whilst those of T_4 and T_3 would be high. This mode of action constitutes a direct effect on the thyroid that is not a result of hormonal imbalance, and is thus regarded as relevant to humans.

Several organisations and committees have published policies or strategies for the interpretation of data on thyroid toxicity, in particular carcinogenesis. It is generally accepted that regulation of thyroid function through the HPT-axis is basically similar in humans and rats, but that there are substantial quantitative interspecies differences in the physiological disturbance of the HPT-axis by non-genotoxic chemicals. The evidence indicates that humans are considerably less sensitive to the development of epithelial follicular thyroid tumours after long-term stimulation than are rats. Furthermore, there are no known chemical human thyroid carcinogens (Hill *et al*., 1989, 1998). RIVM (van Raaij, 2001) suggests that for non-genotoxic chemicals, the following aspects should be addressed before a conclusion of thyroid tumour induction through HPT-axis disturbance is reached:

- *Evidence for a (histo)pathological sequence of events characteristic of prolonged thyroid stimulation*. This includes hypertrophy of follicular cells, increased vascularity, increase in thyroid weight and size, follicular hyperplasia and, eventually, nodular proliferation of follicular cells and neoplasia.
- *Evidence for sustained alterations in circulating hormones*. Elevation in the circulating level of TSH or its turnover (i.e., measurement of TSH levels is a prerequisite for a valid evaluation). Changes in the circulating levels of thyroid hormone or their turnover. As not all substances will affect both T_4 and T_3 levels, the measurement of both hormones is necessary for a valid evaluation. Because of the feedback systems, changes in these hormone levels may only be transient. These alterations in circulating levels are observed relatively quickly after chemical exposure and so can be detected in sub-acute and sub-chronic studies.

- *Information or experimental evidence on the mode of action by which the HPT-axis is disturbed*. The mode of action can be substantiated by general information (e.g., comparison with related chemicals, SARs) or by experimental evidence on the substance (e.g., measurement of liver enzyme induction).

Where the above points are met and the chemical is non-genotoxic, it has been suggested that thyroid follicular tumours induced by such chemicals are not relevant for human carcinogenicity risk. The IPCS has used such an evaluation to assess the carcinogenic potential of thiazopyr in humans (Dellarco *et al*., 2006). It was concluded that, although the postulated mode of action could theoretically operate in humans, marked quantitative differences in the inherent susceptibility for neoplasia to thyroid hormone imbalance in rats meant that thiazopyr did not pose a carcinogenic hazard to humans.

The preceding information relates to the human relevance of thyroid follicular tumours induced in rats that are presumed to be a consequence of HPT-axis disturbance. There is less clarity about the relevance of thyroid effects such as hypertrophy, hyperplasia and nodular lesions to risk assessment. RIVM (van Raaij, 2001) has stated that the disturbance of the HPT-axis itself is toxicologically relevant for humans, since regulation of thyroid function is similar in rats and humans; however, it is acknowledged that the latter are less susceptible than the former. Consequently, disturbance of the HPT-axis is considered by RIVM to be a hazard indicator for humans that should be taken into account when setting NOAELs and health-based reference values. If this effect is the major or critical toxicological endpoint in rats, it is suggested that it might be justifiable to reduce the interspecies assessment factor, based on the lower sensitivity of humans to this effect compared with rats. The USEPA has also adopted the approach of presuming that adverse non-cancer thyroid effects in rodents owing to chemically-induced HPT-axis disruption are relevant to humans (Hill *et al*., 1998). A newly emerging issue, in relation to the current debate about endocrine disruption and the issue of "the critical window of susceptibility", is the lack of consensus as to whether thyroid effects in adults caused by liver enzyme induction should be totally dismissed, especially if no developmental neurotoxicity study has been conducted. Thyroid insufficiency during pregnancy has been associated with developmental neurotoxicity (thyroid hormones are important in brain development). For this effect, there is no evidence for the assumption that rats are more sensitive than humans

Conclusion on thyroid toxicity

The rat is considered to be a highly sensitive model for thyroid effects because it lacks the high-affinity protein binding of thyroid hormones that acts to buffer circulating levels of T_3 and T_4. Consequently, there is an argument that transient changes in circulating rat hormone levels would not be relevant to humans irrespective of the cause. Overall, the prevailing view is that humans are less sensitive to thyroid carcinogenesis than are rats. It is generally accepted that non-genotoxic chemicals that induce thyroid follicular tumours in (especially male) rats through disturbance of the HPT-axis do not merit classification for carcinogenicity, provided that there is sufficient information to conclude that the mode of action is one that is not relevant to humans; in these cases the thyroid tumours would also not be used as the critical effect in a risk assessment. In contrast, non-tumour effects (such as changes in the thyroid morphology, hypertrophy, hyperplasia) mediated by such a disturbance of the HPT-axis are considered by some regulatory authorities to be relevant for risk assessment purposes. In all cases, the onus is on the company to demonstrate that the mode of action is not relevant to humans; otherwise, the default assumption is that the observed effects are relevant and will be taken into account in the risk assessment.

6.5. General organ weight changes

Other than increases in relative liver weight there are no indications that changes (increases or decreases) in relative organ weights could be considered as an adaptive response to xenobiotic administration. There is the potential to demonstrate that increases in relative organ weights are not relevant to human exposures for example due to being secondary to some of the rodent specific effects described above. Therefore the default assumption would be that any statistically significant alteration in relative organ weights is treated as adverse until a scientifically based case is presented to the contrary.

6.6. Inhibition of acetylcholinesterase

One of the few, if not the only toxicological end-point to have a widely agreed adversity cut off is acetylcholinesterase. Discussions at EPCO and PRAPeR meetings have agreed that an inhibition of acetylcholinesterase by 20% or more should be treated as adverse. The JMPR has concluded[31] [32] that statistically significant inhibition of erythrocyte or brain acetylcholinesterase by 20% or more represents a clear toxicological effect and any decision to dismiss such findings should be justified. Statistically significant inhibition of less than 20% or statistically insignificant inhibition above 20% indicate that a more detailed analysis of the data should be undertaken. The toxicological significance of these latter findings should be determined on a case-by-case basis. Considerations affecting such determinations include the shape or slope of the dose-response curve, assay variability, and correlation with clinical signs.

6.7. Clinical chemistry/haematology changes

Based on the previous analyses, clinical chemistry changes are very rarely used in isolation as the basis for setting a NOAEL in a study. They are usually supportive of other, generally pathological, changes. Many clinical chemistry parameters show relatively large variability of 2 or more fold [33][34] making it difficult to set a defined value as a marker of adversity as this would have to be relatively large other wise it would be well within the normal range. For haematology values it is possible to set threshold for certain effects, particularly those relating to erythrocyte parameters. The World Health Organisation has defined haemoglobin levels that are considered to be consistent with anaemia[35]. However, what is less clear is what is the lower level of haemoglobin for normal oxygen carrying function in humans and how does this relate to the typical haemoglobin levels in laboratory animals. The production of methaemoglobin (MetHb) is seen following exposure to a number of synthetic chemicals e.g. substituted anilines. In humans MetHb formation can be indicated by clinical signs such as blue/grey appearance of the extremities. In some humans cyanotic signs can be observed at MetHb levels of below 6% although most individuals can tolerate levels of 10%. Levels of 6% MetHb seems to be a threshold for the occurrence of clinical signs due to MetHb formation in sensitive individuals. as being adverse. Solecki et al (2005) proposed that for acute exposure to methaemoglobin-inducing xenobiotics, a level of about 4% or more above background level in dogs and a statistically significant increase by comparison with controls in rodents is considered to represent a conservative approach to
setting an ARfD. The difference in approach between rats and dogs is due to the much greater ability of rats to reduce MetHb; dogs have a similar capacity to humans.

[31] http://www.who.int/foodsafety/chem/jmpr/en/prst_wp_gls.pdf
[32] http://www.fao.org/ag/AGP/AGPP/Pesticid/JMPR/Download/98_rep/REPORT1998.pdf
[33] http://www.criver.com/SiteCollectionDocuments/rm_rm_r_CD_Rat_clinical_pathology_data.pdf
[34] http://www.criver.com/en-us/newsevents/whatsnew/pages/access_clinical_pathology_data.aspx
[35] http://whqlibdoc.who.int/publications/2008/9789241596657_eng.pdf

6.8. Overall conclusion on non-adverse / adaptive findings

Although there are a number of modes of action that support a contention that findings in laboratory animals are not of relevance to humans there are very few cases where these can be turned into generic cut-off levels for adversity. Based on general acceptability it could be possible to set thresholds for non-adverse changes for a small number of parameters:

- Changes in body weight gain of <10%;
- Increased relative liver weight of <10%;
- Inhibition of acetylcholinesterase of <20%;
- Increases of MetHb levels of <4% in dogs.

For all other effects any statistically significant changes or changes outside the historical control range should be considered as adverse and relevant to humans until a scientific case is presented to the contrary

7. EFSA TASK 6:

Collect and scrutinize relevant scientific literature or other relevant publications or information sources (e.g. guidance documents or government reports) on the appropriateness of the current approaches to deal with the uncertainty in the establishment of health-based guidance values and, if relevant, propose other science-based approaches of application and refinement of uncertainty factors in setting the ADI, AOEL or "AAOEL" values.

Literature searches were performed by information scientists using the search criteria described in Annex 5. Based on the abstracts obtained, potentially relevant references were identified by the toxicology specialists, obtained and evaluated. The toxicologists also performed targeted online searches and obtained copies of references cited in core texts and primary references.

This section addresses methodology to address the uncertainty surrounding the derivation of health based guidance values and the description of these uncertainties.

7.1. Default Uncertainty Factors

In general the risk assessment of chemicals for threshold effects on human health involves a comparison of the hazardous properties of the chemical with a measure or estimate of human exposure. In this approach, the first step is normally the identification of a critical effect (or effects) and the experimental dose level at which that effect is not detected (NOAEL), or at which the effect is found to be minimal in incidence and severity (LOAEL). These dose levels may be identified from studies in human populations, but, in the vast majority of cases, reliance must be placed on data from studies in experimental animals and other test systems. In some cases, a mathematically-derived value, the benchmark dose (BMD), may be used as an alternative to the NOAEL or LOAEL. Whichever is available or chosen, this starting point is then used to derive an exposure standard or reference value considered to represent a level of exposure or intake at which it is believed there is little, if any, likelihood of developing ill-health effects. This reference value is then compared directly with the measure or estimate of exposure.

Such methodology for the establishment of reference values has to deal with many uncertainties in terms of the available toxicological information. These uncertainties generally include the need to extrapolate between species (i.e. from an experimental animal species to humans) and the need to account for variability in the potentially exposed human population, but may also include uncertainties owing to limitations in the database (e.g. no long-term studies, not a full exploration of the range of potential toxic properties). Not all of these uncertainties may be encountered in all situations but it is very rarely the case that sufficient information from human experience will be available to address all the uncertainties that arise in a risk assessment. Thus, over the years, approaches have been developed to address these uncertainties in a systematic and generally consistent manner. The most widely used and widely accepted approach has involved the application of factors, referred to as uncertainty factors (or safety factors or assessment factors), to allow for these uncertainties. These are applied either directly to the NOAEL or LOAEL (or BMD or BMDL) in the derivation of reference values or as a framework against which to judge the adequacy of a derived hazard or exposure ratio (margin of exposure).

When knowledge of the hazardous properties of a substance is at a basic level (e.g. a basic set of experimental information, which would include long-term studies in animals and information on reproduction and development but little on toxicokinetics, modes or mechanisms of action or knowledge of human variability), then the approach in many areas of regulatory decision-making has been to apply default uncertainty factors. These have usually been of the numerical order of 10 each

to allow for uncertainties in interspecies extrapolation and intraspecies variability, and have been used in order to generate a position deemed to offer the required degree of confidence in health protection being sought. These default values assume that the average human is 10 times more sensitive than the average animal and the most sensitive human is 10 times more sensitive than the average human. These two default values are then multiplied together to produce an overall uncertainty factor of 100 based on the assumption that interspecies and intraspecies variability affect the value of the reference dose/exposure independently of each other. Such overall uncertainty factor of 100 has been and is still used extensively in assessing the risks to human health posed by chemicals used in different situations, for example, exposures arising from use in consumer products, from food (e.g. pesticide residues), or from exposures arising from environmental pollution in air, soil, water or waste. In relation to pesticides, as indicated in the text under *Activity 1*, of the 224 EFSA conclusions considered, 216 (96.4%) and 222 (99%) had an ADI or AOEL, respectively, derived by the application of an overall assessment factor of 100 or higher.

This approach was originally adopted in the USA in the 1950s, but has since become a well-established international practice used widely in many regulatory contexts and in many fora. These factors are intended to provide a level of reassurance of safety from the harmful effects of exposure to chemicals in the face of limited information; more information would help the risk assessor to make a more accurate prediction of the true level of risk. Although the original exact derivation of these default factors of 10 is somewhat uncertain, many consider that they were based on very limited evidence and arguably had little scientific basis. However, over the years since their introduction, and particularly since the 1980s, there have been an increasing number of scientific analyses presented on various aspects of these factors, such that in general they are now supported scientifically as providing a default position deemed to match the degree of reassurance sought when information is limited.

For example, there have been numerous reviews of the appropriateness of the 10-fold factor for human variability, based on human variability in kinetics and dynamics. Attempts have been made to quantify what proportion of the human population would be protected by these safety factors. The analyses of Renwick and Lazarus (1998) indicated that the 10-fold uncertainty factor for human variability would cover the vast majority of exposed individuals assuming a normal or a log-normal distribution (such models would not cover 100% of a population unless the uncertainty factor is infinity). The data on therapeutic drugs indicate that the current uncertainty factor is a reasonable default value and would cover the normal human population to greater than 99%. However, the usual 10-fold factor would not allow adequately for human variability when there is a genetic polymorphism in the main route of elimination (Dorne, 2012). Also, differences between healthy adults and some subgroups of the population (such as preterm infants) may not be covered adequately. Differences between human neonates and human adults would not need to be allowed for by the uncertainty factor for human variability if there were a developmental study in neonatal animals, because the interspecies comparison would take into account any risk related to immaturity, providing that the neonatal rat was at least as immature as the neonatal human.

Recent years have seen a considerable move forwards in the development of technologies and generation of data that may help to improve our approaches to dealing with toxicological uncertainties. The replacement of default uncertainty factors in risk assessment may be possible where chemical-specific information can be used in frameworks designed to accommodate such information, such as that developed internationally under the auspices of the International Programme on Chemical Safety (IPCS) for the use of chemical-specific adjustment factors. Such information can also be used in more advanced approaches such as physiologically-based pharmaco-kinetic (PBPK) modelling, reducing the need for default uncertainty factors in at least some aspects of risk assessment. Probabilistic approaches may be useful either as an alternative to simple multiplication in combining a number of uncertainty factors, or as the basis for an approach that differs from the traditional risk assessment paradigm.

Additional uncertainty/safety/assessment factors

As well as the defaults used to allow for the uncertainties in interspecies extrapolation and intraspecies variability, other default factors have been developed and used for situations where other uncertainties arise because of deficiencies in the available database, for example the lack of a long-term study, the use of a LOAEL instead of a NOAEL, data gaps or the need to make allowance for especially severe or irreversible effects. In general, default uncertainty factors of up to 10 have been used in such situations. Alternatively, analyses of aggregated existing experimental data have been used to derive more specific factors (e.g. when comparing the ratio of experimental NOAELs over varying exposure periods), but some of these analyses have fundamental flaws as they are reflective of, for example, dose selection rather than any true factor that reflects the relationship between the values being compared. Although the use of defaults or database-derived factors represent a standardised and consistent approach, the use of expert judgement in such situations is also very important.

It has been noted that an additional severity factor of up to 10 is generally used in some fora when deriving a reference value on the basis of a severe effect (e.g. teratogenicity). Many agree that there is no scientific basis for the application of such factor, especially if a robust NOAEL is the basis for the derivation of the reference value. The application of the severity factor and its magnitude should therefore be a risk management decision justified by the requirement of a higher level of protection for situations where exceptionally severe effects may occur.

Similarly, another additional factor occasionally applied when deriving a reference value is a figure to account for a steep dose-response relationship. Again, many agree that there is no scientific basis for the application of such a factor because when the dose-response is steep, there is more confidence and certainty in the identified NOAEL and that effects are unlikely at exposures below it.

7.2. Allometric scaling

Human risk assessments are often based on toxicity data from laboratory animal species, thereby necessitating a number of extrapolations to estimate the exposure conditions for which a similar toxicity would occur in humans. A critical step in this process is relating the exposure-dose-response relationships for laboratory animals to those pertaining to humans; that is, the need to adjust the exposure used in an animal study to a human “equivalent” exposure. The most scientifically sound approach by which this may be accomplished is through the use of a physiologically-based toxicokinetic (PBTK) model of site-specific dosimetry. An intermediate approach is the use of information on species differences and chemical-specific toxicokinetic and toxicodynamic data that enable the derivation of cross-species adjustment (for example, chemical-specific adjustment factors). In most cases, however, there are insufficient toxicokinetic/dynamic data available to compare internal doses between different species; in these cases, default approaches are adopted to enable a risk assessment to be performed.

When extrapolating data from animal studies to humans, account should be taken of species-specific (interspecies) differences. Interspecies differences result from variation in the sensitivity of species owing to differences in toxicokinetics and toxicodynamics. Historically, a default interspecies assessment factor of 10 has been applied to extrapolate from the average animal studied to an average human being, assuming humans might be 10-fold more sensitive than experimental animals. Subsequently, it was suggested that the interspecies factor of 10 could be subdivided into toxicokinetic (4.0) and toxicodynamic (2.5) components (e.g. Renwick, 1993).

The use of experimental mammals such as rats, mice and dogs to predict adverse health effects in humans is premised upon the high degree of physiological, biochemical and anatomical similarity among mammalian species. The historical default approach to the extrapolation of data from laboratory animals to humans therefore assumed a body weight (BW) scaling from animals to humans of 1 ($BW^{1/1}$), with a full uncertainty factor of 10 for interspecies differences, regardless of the animal species used. However, even if fundamental similarity is presumed, one must allow for the fact that the greater size and lifespan of the human relative to the experimental animal has a significant impact on the amount of chemical intake needed to provoke a given response.

In comparison with rodent species, the administration of the same quantities of harmful substances per kg body weight (kg bw) frequently results in more severe effects among larger animals and man (Kalberlah and Schneider, 1998). This implies that a scaling method with dose proportional to a power of body weight less than 1 would be appropriate for determining equally toxic doses. This is because body weight does not correlate with many physiological functions of the body such as basal metabolic rate or caloric demand (oxygen consumption), which are important determinants of kinetic parameters that affect the toxicity of a chemical. However, the analyses described under activity 1 show that rats are used more regularly than dogs in setting reference values, which would tend to contradict this argumentation relating to body size and sensitivity to chemicals – assuming dose spacing, dosing methods and levels of investigation are not confounding the comparison.

To enable the refinement of risk assessments, a theoretical approach called allometry has been developed that aims to characterise the impact of scale on a chemical's toxicological potency across species; i.e., how the regular patterns of size, physiological pace and lifespan across the spectrum of differently sized mammals affect the toxicological process. Allometry has been proposed as a general default procedure to extrapolate toxicologically equivalent doses of inhaled and orally-administered chemicals from laboratory animals to humans.

A large number of characteristics and functions of mammalian biological systems have been examined for their relationship with BW. Volumes and capacities tend to retain their proportionality across species, i.e., they show scaling of $BW^{1/1}$, since BW increases in direct proportion to blood volumes and organ weights. In contrast, a number of physiological processes increase in absolute values but in proportion only to the 0.75 power of the body weight ($BW^{3/4}$). Allometry therefore proposes that quantitative differences across mammalian species in physiological processes can be seen largely as the consequence of fundamentally similar anatomical and biochemical machinery operating at different rates in differently sized species, with smaller species having faster physiological "clocks". A mouse, for example, is carrying out approximately the same set of physiological processes as a human, but each proceeds at a rate that tends to be approximately 7-times faster. In theory, the various processes stay in proportion to one another, but all of them are relatively accelerated in smaller species. For those processes that involve rates and time, a decrease in the absolute value may occur; for example, although the body mass and absolute heart mass are both about 2300-fold greater in humans than in mice (scaling to $BW^{1/1}$), and cardiac output is about 300-fold greater than in mice (scaling to $BW^{3/4}$), the heart rate in humans is approximately 7-fold less than in mice (scaling to $BW^{-1/4}$).

The physiological processes that scale to $BW^{3/4}$ are those that are driven by the caloric demand (basal metabolic rate) such as cardiac output, energy utilisation (glucose turnover), blood flow, perfusion of liver and kidneys, glomerular filtration, minute volume and the maximum velocity of metabolic pathways with saturable metabolism. As these physiological processes are important determinants of key kinetic parameters, such as elimination/clearance, which in turn affect the toxicity of a chemical, it can be predicted that, generally, toxicokinetic differences between species are accounted for by scaling the administered dose levels with $BW^{3/4}$. Hence, allometric scaling on the basis of caloric

demand extrapolates doses according to an overall assumption that equitoxic doses (when expressed in mg/kg bw/day) scale with body weight to the power of 0.75. This results in different default allometric scaling factors for the different animal species (see table below) when compared with humans (i.e., it addresses predominantly the toxicokinetic and some of the toxicodynamic aspects of the inter-species uncertainty factor).

Many of the same empirical investigations that support $BW^{3/4}$ scaling also examined whether scaling on the basis of surface area (i.e. to $BW^{2/3}$) was an appropriate method to determine equitoxic doses across species. [Surface area scaling is equivalent to scaling doses by the 2/3 (0.67) power of the BW because the surface area of similarly shaped objects varies as the 2/3 power of their volume.] Although surface area scaling was not completely rejected by the experimental data, there is no explicit rationale based on allometric variation of the underlying anatomy, physiology and metabolic size of different mammalian species to support this approach (Rhomberg and Lewandowski, 2004). The differences between the area or weight scaling are generally small (<2 fold), as shown in table 7 below.

Table 7 - Allometric scaling factors for different species as compared with humans[a]

Species	Body weight (kg)	Allometric scaling factor	
		weight	area
Rat	0.250	4	6
Mouse	0.025	7	13
Hamster	0.11	5	-
Guinea pig	0.8	3	-
Rabbit	2	2.4	-
Monkey	4	2	-
Dog	18	1.4	1.5

[a] Assuming the human body weight is 70 kg

The factors are derived according to the formula:

$$\frac{bw_{human}/bw_{animal}}{(bw_{human}/bw_{animal})^{0.75}} = (bw_{human}/bw_{animal})^{0.25}$$

The use of these allometric scaling factors results in the derivation of a human equivalent exposure, specifically, a human equivalent dose in the case of oral exposure. Compared with the application of the default interspecies factor of 10, the use of a combination of species-specific allometric scaling factors with a default factor for remaining uncertainties in interspecies differences results in overall interspecies factors which range from 3.5-4.2 (dog) to 17.5-21 (mouse), depending on the animal species from which the point of departure has been identified.

It should be noted that these species-specific allometric scaling factors are average values applicable to average individuals of average body weight. Therefore, particular caution should be exercised when applying such factors to a toxicological reference point derived from a study using young adult animals with relatively small body weights (e.g. a 28-day study in rats) to establish a chronic reference value applicable to a population of adults whose body weights are much larger.

The allometric assumption that important kinetic processes and related toxicological effects are driven by the basal metabolic rate ($BW^{3/4}$), which was originally predicted mathematically, has been subsequently substantiated by a considerable number of empirical investigations which have examined the kinetics and toxicology of different groups of chemicals in relation to body weight in a range of different animal species as well as humans (Dedrick et al., 1970, 1973; Walton et al., 2001a,

2004; Schneider et al., 2004). These studies suggest that allometric scaling on the basis of caloric demand or basal metabolic rate, i.e. to the ¾ power of the body weight, represents a reasonably well-supported approach for extrapolating oral doses across species, particularly at relatively low doses at which metabolism is not saturated and clearance rates are pseudo-first order. The concordance between the hypothesis and experimental observations lends credence to the theory of $BW^{3/4}$ scaling. There is also empirical evidence to indicate that allometric scaling on the basis of caloric demand is generally applicable to substances that are essentially renally excreted, but not to compounds that are highly extracted by the liver and excreted in the bile; it appears that species differences in biliary excretion and glucuronidation are independent of caloric demand (Walton et al., 2001b).

Several considerations with regard to assumptions and limitations are pertinent to the use of $BW^{3/4}$ as a default for the estimation of toxicologically equivalent doses, as summarised below:

- **Metabolism and clearance.** Allometric scaling according to basal metabolic rate would apply most appropriately to those substances for which the unmetabolised parent or a stable metabolite is the relevant, systemically available toxic species and clearance is according to first-order processes. Conversely, the applicability of allometric scaling when toxicity is a consequence of exposure to a very reactive parent compound (or metabolite) that is not removed from the site of formation (e.g. local toxicity at the port-of-entry), is less well supported (USEPA, 1992).
- **Measure of delivered dose: choice of the appropriate dose metric.** The use of $BW^{3/4}$ scaling is most appropriate for substances where the measure of dose associated with the toxic effect is the area under the curve (AUC) rather than the C_{max} (USEPA, 2006). This is because the normalisation of dose across species, based on the concept of physiological time scaling, is in terms of exposure to a concentration over a duration of time: the measure of dose that is being scaled for kinetic equivalency between species in this process is the AUC.
- **Early life stages.** Some reports have indicated that allometric scaling may not be appropriate when deriving reference doses specific for children, owing to disproportionate development of biochemical and physiological processes and differing allometric patterns amongst various sized individuals of the same species. However, recent work shows that $BW^{3/4}$ scaling is descriptive of toxicokinetic differences among ages, including very early life stages, down to about 2 months (USEPA, 2006). Furthermore, it should not be forgotten that even when the "target tissue" of a toxicant is the foetus or the developing pup, such as in developmental or multigenerational reproductive studies, the doses requiring scaling have been administered to adult animals and require extrapolation to adult humans. The USEPA has noted that the application of $BW^{3/4}$ scaling from another species by scaling to the body weight of children rather than adults would yield a higher equivalent dose. It concludes that scaling to children's body weights might not be appropriate for RfD or short-term guidance values intended to apply to a population that includes young infants and children, because of the slower clearance during this life stage and the limited toxicokinetic data available in early life (USEPA, 2006).
- **Toxicokinetics and toxicodynamics in toxicological equivalence.** Species differences in dose-response functions may be elicited as both a consequence of distribution of agent affecting the target-tissue dose between species (toxicokinetics); and, from intrinsic differences in the tissue response between species. Achieving toxicological equivalence across species requires that aspects of both toxicokinetics and toxicodynamics be considered, and hence the use of $BW^{3/4}$ to achieve toxicological equivalence for interspecies differences implies that scaling is inclusive of both these aspects. Although many physiological processes relating to kinetics conform to the $BW^{3/4}$ relationship, it should not be concluded that this scaling factor encompasses all kinetic factors related to toxicity. Also, it is not intended that $BW^{3/4}$ scaling does not address any toxicodynamics: some toxicodynamic processes, e.g. cellular repair and regeneration, signalling cascades and proliferative response, also scale as a fractional power of BW. Therefore, although

$BW^{3/4}$ scaling predominantly addresses factors involved in estimating toxicokinetics, it also addresses some toxicodynamic factors.

- **Acute exposures.** There is evidence to suggest that allometric scaling is not appropriate for acute lethal effects (Rhomberg & Wolff, 1998) as these effects, which are accomplished by an immediate and intolerable level of damage to some critical homeostatic processes, may be independent of caloric demand and related physiological processes which affect toxicity. However, for acute exposures involving other, less severe effects, in which the operative physiological processes are comparable to those of chronic exposures, $BW^{3/4}$ scaling is considered a reasonable approach (USEPA, 2006).
- **Portal-of-entry issues for oral exposure.** It is to be noted that allometric scaling should not be applied if the effects are not dependent on basal metabolic rate or systemic absorption, e.g. in the case of local effects. In general, as long as route-to- route extrapolation is not needed, allometric scaling should also not be applied in cases where doses in experimental animal studies are expressed as concentrations (e.g., in mg/m^3 in air, ppm in diet, or mg/l in the drinking water): these are assumed to be already scaled according to the allometric principle, since ventilation rate and food intake directly depend on the basal metabolic rate. However, once the concentration (e.g., ppm in diet) has been converted into a dose (e.g., mg/kg/day), an allometric scaling factor can be used. Thus, it is the dose unit (original or transformed), and not the (experimental) route of application, that triggers the necessity for a species-specific factor for allometric scaling.

In applying allometric scaling, it is important to recognise that it is the absolute intake or exposure in the experimental animal, in mg, that is scaled to the human, not, for example, mg/kg (which is actually a rate process). This results in an absolute, scaled human intake that can then be converted to a mg/kg value. Guidance on how to convert an intake or dose in a laboratory animal to a human equivalent exposure is given in appendices A and B of USEPA (2006). Table 8 below demonstrates the $BW^{3/4}$ scaled human equivalent exposures converted from a 10 mg/kg exposure in different species (which, if applying the historical $BW^{1/1}$ factor, would lead to an estimated human exposure of 10 mg/kg).

Table 8 - Estimation of oral exposure in humans[a] based on $BW^{3/4}$ scaling of a 10 mg/kg exposure in rats, mice and dogs

Species	Absolute animal intake	BW(h)/BW(a)	BW scaling factor	BW scaled human intake & dose
Mouse	0.25mg/0.025kg	70/0.025=2800	$2800^{3/4}$=385	385 x 0.25mg = 96mg 96mg/70kg = 1.4mg/kg
Rat	2.5mg/0.25kg	70/0.25=280	$280^{3/4}$=68	68 x 2.5mg = 170mg 170mg/70kg = 2.4mg/kg
Dog	120mg/12kg	70/12=5.8	$5.8^{3/4}$=3.7	3.7 x 120mg = 444mg 444mg/70kg = 6.4mg/kg

[a]Assuming a human body weight of 70 kg

The use of $BW^{3/4}$ scaling in combination with a reduced default interspecies uncertainty factor is being applied by an increasing number of regulatory bodies/agencies (e.g. USEPA, FDA, TNO, BAUA, ECHA) around the world when establishing health-based reference doses. The additional default factor normally ranges from 2.5 to 3 and takes into account the remaining interspecies differences, in particular toxicodynamic differences. This should lead to more refined and scientifically-based risk assessments: the historical default interspecies factor of 10 is arbitrary (although some post-hoc work by e.g. Renwick *et al.* have shown that in the vast majority of cases a factor of 10 is protective), whereas species-specific allometric scaling factors are supported by some

experimental data and by fundamental physiological and biochemical processes that underlie the kinetics of toxicants.

The answers to the questionnaire indicate that the respondents do not routinely use allometric scaling in the derivation of reference doses (see questionnaire results, Task 3). One respondent to the questionnaire had previously used scaling but was no longer using it as it was not now accepted within the EU system for pesticides. However, allometric scaling is an integral part of the risk assessment process in one of the major EU chemical regulatory schemes (REACH; allometric scaling is used in the derivation of DNELs (derived no-effect levels) for threshold effects, see task 3) and so it is becoming more widely used by industry and European regulatory authorities. The USEPA[36] is planning to use allometric scaling in the future.

7.3. PBPK Modelling

A physiologically based pharmacokinetic (PBPK) model is a structural mathematical model, comprising the tissues and organs of the body, each perfused by, and connected via, the blood circulatory system. The principal application of PBPK models is in the prediction of the *target tissue dose* of the parent chemical or its reactive metabolites. PBPK modelling provides a mechanistic approach to both understanding the temporal behaviour of compounds within the body and predicting what is likely to happen in plasma and tissues over a wide range of conditions.

Use of the target tissue dose of the toxic moiety of a chemical in risk assessment provides a better basis for relating the exposure to the observed toxic effects than does the external or administered exposure concentration of the parent chemical. Prediction of target tissue dose following different exposure scenarios, routes, doses and species can help reduce the uncertainty associated with conventional extrapolation approaches. The mechanistic and biological plausibility of the models is the basis for associating greater confidence to such extrapolations.

Parameters in PBPK modelling
PBPK models comprise four main types of parameters:

1. Physiological
2. Anatomical
3. Biochemical
4. Physicochemical

Physiological and anatomical parameters include tissue masses and blood perfusion rates, estimates of cardiac output and alveolar ventilation rates. Biochemical parameters include e.g. enzyme metabolic rates and polymorphisms, enzyme synthesis and inactivation rates, receptor and protein binding constants. Physicochemical parameters refer to e.g. partition coefficients, vapour pressures, solubilities in different media. A partition coefficient is a ratio of the solubility of a chemical in a biological medium, usually blood-air and tissue-blood.

Anatomical and physiological parameters are readily available and many have been obtained by measurement. Biochemical and physicochemical parameters are compound specific. When such parameters are measured and used to construct an *a priori* model that qualitatively describes a dataset,

[36] http://www.epa.gov/raf/publications/interspecies-extrapolation.htm

then confidence in such a model should be high. In the absence of measured data, such as partition coefficients, these may be estimated with tissue composition-based algorithms (Poulin and Krishnan 1995, 1996; Theil *et al.* 2003). Metabolic rate constants may be fitted with a PBPK model, although this practice should only be undertaken if there are no other alternatives.

The importance of any single or set of parameters within a model should be determined by applying sensitivity analysis. Sensitivity analysis is a means of evaluating how sensitive the model output is to any perturbation of any single or set of parameters. Therefore, confidence in a model that contains estimated parameters may still be high, if those parameters do not significantly influence model output. Conversely, influential estimated rather than measured parameters would reduce confidence in a model.

As knowledge of the physiological, anatomical and biochemical system and how compounds interact with it increases, so will the ability of the PBPK model to predict the likely behaviour of compounds from relatively limited data on the compound. To provide meaningful predictions, however, it is important to incorporate biological variability and methodological uncertainty in parameter values throughout the modelling process. Unlike with empirical models, compound-specific data from various sources, *in silico*, *in vitro*, and *in vivo*, can be readily incorporated into PBPK models. However, it is critical to verify at every opportunity the quality and utility of the input data against events of interest *in vivo*. The PBPK model approach is flexible in the sense that it has the potential to be continuously updated in the light of new information, whether physiologic, disease, or compound related.

Confidence in PBPK modelling
When used for risk assessment purposes, confidence in PBPK models should be high (Barton *et al*., 2007). Therefore, their predictive capacity should be carefully evaluated with respect to the following aspects (USEPA, 2006):

- model verification (i.e. biological plausibility of the model structure and parameter and correctness of the mathematical equations);
- model validation (i.e. ability of the model to predict the kinetic behaviour of a compound);
- model documentation;
- sensitivity, variability and uncertainty analyses.

When PBPK modelling is used for risk assessment purposes, the whole process of PBPK modelling (i.e. the generation of the model, application of the model, validation of the model, confidence in the model, etc.) should be as transparent as possible. Confidence in the model should be as high as possible. Furthermore, risk assessors, who are using these models, should be able to adequately interpret them and their output. However, there is as yet no consensus as to the means or extent of validation required before a PBPK model can be used for regulatory purposes.

PBPK modelling may have multiple applications in risk assessment by contributing to reduce the uncertainty in its various extrapolation procedures, i.e. between species, individuals, high to low doses, routes and different exposure scenarios.

Interspecies Extrapolation
Interspecies extrapolation of the pharmacokinetic behaviour of a chemical requires quantitative estimates of species differences in the values of these parameters. Tissue-blood partition coefficients of chemicals appear to be relatively constant across species, while blood-air partition coefficients show some species-dependent variability (Gargas *et al.* 1989). Physiological and anatomical

parameters generally vary coherently across species. These parameters are readily available in the literature (Brown *et al.* 1997; ICRP 1975) and can therefore be used in a PBPK model where quantitative differences can be evaluated. The kinetic constants for metabolizing enzymes do not necessarily follow any type of readily predictable pattern (Dedrick and Bischoff 1980). The approach adopted in the past, and one that is still often used, is to apply the "metabolic rate scaling". Therefore, a metabolic rate constant, such as Vmax, obtained in a rodent would be multiplied by body weight of the human raised to the ¾ power to obtain the human whole body equivalent. Whereas a metabolic rate constant estimated in this way may be used in a PBPK model, it is preferable, where possible, to determine such parameters *in vitro* using tissue subcellular fractions or to estimate them by fitting a PBPK model to an appropriate dataset. Furthermore, if a PBPK model is used to extrapolate from animals to humans, the proposed model should be validated by data from humans if these are available, and extrapolations from the model should be within or close to the range of experimental measurements used to validate the model. If there is no validation of the model by data from humans, PBPK models may be used to support an interpretation of toxicological findings rather than as a basis for the derivation of a chemical-specific interspecies kinetic factor.

Intraspecies Variability

Differences in sensitivity to exposure to chemicals within the same species occur as a result of variation in anatomical, physiological and biochemical parameters with age, gender, genetic predisposition and health status. These may be further confounded by nutritional and other lifestyle and environmental factors. The quantification of these parameters using PBPK models is analogous to the quantification of interspecies variability. For example, age-specific parameters would be required to estimate the tissue doses in adults and young children. Such data are increasingly available. The propagation of uncertainty and variability from model parameters to model output can be quantified using probabilistic techniques such as Monte Carlo sampling. A PBPK model is run with parameter values sampled from distributions that reflect the observed variation in each pharmacokinetic parameter in the human population. Each time the model is run with a sampled set of parameter values, effectively representing a single hypothetical human being, the appropriate dose metric for the toxicity of interest is estimated. The process is repeated a large number of times to generate a distribution of the dose metric for a simulated population. It is important to note that human physiological data have a range of values. Therefore, modelling should be preferentially performed with ranges of values leading to distributions of outcome.

High-Dose-Low-Dose Extrapolation

The non-linear kinetic behaviour of chemicals in a biological organism is the result of a number of mechanisms e.g., saturable metabolism, enzyme induction, enzyme inactivation and depletion of glutathione and other cofactor reserves. High-dose-low-dose extrapolation of tissue dose is accomplished with PBPK modelling by accounting for such mechanisms (Clewell III and Andersen, 1987).

Route-to-Route Extrapolation

Route-to-route extrapolations can be conducted quite readily with PBPK models. For example, the procedure would involve describing a model for the inhalation route. Ideally, the model would be validated against an appropriate dataset. Equations describing other routes of administration, such as dermal and oral, may be added later. Again, ideally, the model should be validated against a different, but appropriate dataset for the additional route. In the case of oral uptake, first-pass metabolism and enterohepatic circulation may also be included if significant elimination of parent chemical occurs as a result of these mechanisms (Clewell III and Andersen, 1987).

Conclusion on PBPK

PBPK models will not remove all of the uncertainty from the risk assessment process. The rationale for using PBPK models in risk assessment is that they provide a documentable, scientifically

defensible means of bridging the gap between animal bioassays and human risk estimates. In particular, they shift the risk assessment from the administered dose to a dose more closely associated with the toxic effect by explicitly describing their relationships as a function of dose, species, route and exposure scenario. The increased complexity and data demands of PBPK models must be counter-balanced by the increased accuracy, biological plausibility and scientific justifiability of any risk assessment using them. It follows from this that PBPK models are more likely to be used for data-rich substances where acceptable risks have not been shown using standard default approaches. A guidance document on "Good Practice in PBPK modelling" has been issued by WHO/IPCS (2011) and should be taken into account when PBPK modelling is used in risk assessment.

Notwithstanding the benefits, there are many obstacles to the wider use of PBPK modelling, including a lack of user-friendly modelling software, lack of appropriate and easily accessible relevant physiological and related databases, lack of adequately trained researchers in such modelling and the need for more and increasingly complex data.

7.4. Chemical Specific Adjustment Factors (CSAF)

The use of chemical (compound) specific adjustment (assessment) factors follows on from the work of Renwick and co-workers in sub-dividing the default 100-fold safety factor traditionally used in the derivation of ADIs into four components. The four factors were for interspecies toxicokinetics (4.0), interspecies toxicodynamics (2.5); intraspecies toxicokinetics (3.16), and intraspecies toxicodynamics (3.16). Related to this sub-division was the evidence that the 100-fold factor was protective of human exposures to chemicals. Additional work on therapeutic drugs (Renwick and Lazarus, 1998) indicated that in some circumstances the default factors were not appropriate and that specific factors (higher or lower) could be applied if supported by appropriate data. This was utilised by the IPCS harmonisation project which produced a guidance document on the use of specific adjustment factors in dose-response assessments (IPCS, 2005).

In order to support the movement from one of the default factors, some specific data need to be available that demonstrate that one or more of the default factors are not appropriate for a particular chemical. A framework for this was presented by the IPCS (IPCS, 2005).

IPCS notes that if using reduced overall factor (i.e. <100) for a particular end-point then an evaluation should be performed to confirm that alternative end-points, that would remain subject to the 100-fold overall factor, would not become the lead effects for the risk assessment. Recent work by Dorne (2010) has indicated that for certain metabolic pathways with human polymorphic variation, the intraspecies toxicokinetic factor of 3.16 was inadequate to cover the full range of human variation. It was acknowledged by Dorne that the polymorphic variation might not be relevant at low dose levels typical of human environmental exposure levels. Such polymorphic variations could also have an impact on the ability to metabolise natural components of the diet.

Application of CSAFs

CSAFs have been used in the derivation of reference doses for pesticides by the WHO JMPR. The underlying reasoning has not always been precisely described, but the following approaches have been used.

- Reduction of the interspecies dynamic factor by a factor of 2 for abamectin owing to the use of the NOAEL for a particularly sensitive sub-group of animals (neonatal rats) (JMPR, 1997).

- A reduction of the interspecies factor by 2 for tolylfluanid (JMPR, 2002) owing to the low species variation in fluoride deposition in teeth and bones.
- A reduction in the toxicokinetic factors by 2 to give overall factors of 25 for ARfDs and ADIs from animal data or 5 for human data in deriving reference doses for N-methyl carbamate, cholinesterase inhibitors (JMPR, 2000 Annex 5; JMPR, 2002 report item 2.2). The reduction is because these compounds have a critical effect (acetylcholinesterase inhibition) that is dependent on Cmax and is rapidly reversible, requiring no metabolic activation, and being rapidly detoxified by a single metabolic reaction.

An overall factor of 25 for the derivation of the ADI and ARfD for methamidophos (JMPR, 2002) based on data that showed negligible species differences in the pattern of cholinesterase inhibition in rats, dogs and humans and an effect that was Cmax dependent

CSAFs have also been considered by EFSA (e.g. the evaluation of caramel food colours[37]) and JECFA in the consideration of contaminants such as dioxins, methylmercury (IPCS, 2005) and cadmium[38]

Data required to support a CSAF

The IPCS framework sets out options for generating data to support a CSAF but as the final decision relies on expert judgement it was not possible to give definitive conclusions on what level of information is required to support the use of a CSAF. The data requirements vary with each particular factor and include in vitro dynamic data (e.g for cholinesterase inhibition or receptor binding), kinetic comparisons either using either human volunteers or PBPK modelling, etc. Recently developed techniques such as genomics or metabolomics and the availability of human cells and liver slices could be used to provide information relevant to variations in the human population / sensitive sub-groups or variations in response between humans and animals. Revisions to the data requirements for pesticides in the EU have included data on metabolism in test species and humans which could be of value in the derivation of CSAFs.

One of the aspects identified in the IPCS framework is that the kinetic data should be appropriate to the dose levels applicable to each species, i.e. in test animals the data should relate to the NOAEL / LOAEL range and not be subject to any non-linear kinetics and the comparison to humans should be at the ADI / ARfD dose range. The data should relate to the dose of the toxicologically active molecule or molecules (parent compound and / or metabolites) at the target tissue or receptor, e.g. similar plasma levels would not necessarily relate to equal doses to the fetus or brain that are protected by the placenta or blood brain barrier, respectively.

For toxicodynamic data, these should be generated in such a way that any effects of kinetic variation are corrected for.

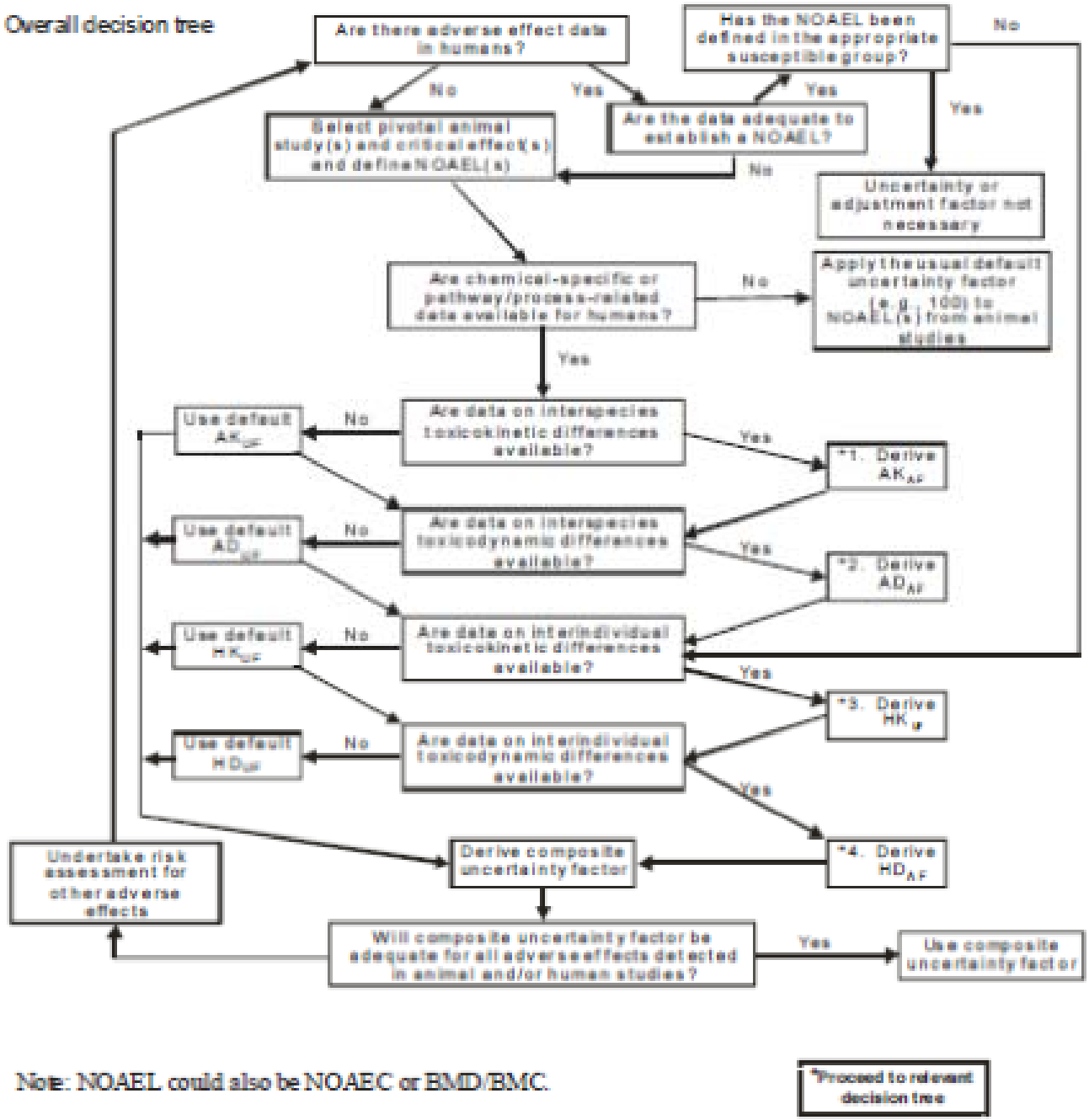
For many chemicals, particularly those with a poorly understood mode of action, the available information will not be adequate to support the derivation of CSAFs. The use of CSAFs resulting in an overall factor of <100 is precluded by the current EU scheme for pesticides.

Conclusion on CSAFs

The principles behind CSAFs have been available for over a decade, yet appear to be unused other than by JMPR for pesticides and by some other committees for contaminants. The reasons behind this

[37] http://www.efsa.europa.eu/en/efsajournal/pub/2004.htm
[38] http://whqlibdoc.who.int/publications/2011/9789241660648_eng.pdf

are unclear. A number of techniques are available to investigate the kinetic and dynamic variation between species and between human sub-groups that would permit the refinement of reference values. There appears to be a reluctance among companies producing chemicals to generate appropriate data to submit to regulatory authorities.



CSAFs are a tool that permit the refinement of a health based guidance values by the generation of data to show that the default inter and intra-species kinetic factors are not appropriate to a particular risk assessment. If adequate data are available (as outlined by IPCS, 2005) they should be taken into account in the overall risk assessment.

Low dose linear extrapolation and MoE approach for cancer and noncancer endpoints

The process of conducting risk assessments for chemical agents has evolved over the past century and has increasingly become more formalised. The United States Environmental Protection Agency (USEPA) has probably had greater impact than any other agency on a global scale in fostering the use of risk assessment. The National Research Council (NRC, 1983) has also offered advice, documented in numerous NRC reports, on the risk assessment process and how it should be used.

Prior to the 1970s an exposure guideline based on animal data was set by applying safety factors (SFs) to a NOAEL. The NOAEL was interpreted as an estimate of the dose where a threshold for adversity could be expected in the test animals, and the safety factors were applied to account for differences in sensitivity between the test animals and humans, and for human heterogeneity. During the 1970s it was realised that there might not be a risk-free exposure to chemicals that could initiate cancer by causing a mutation in a single cell (i.e. genotoxic carcinogens) (NRC, 1977). As a result, risk assessment began to incorporate the assumption that no amount of exposure to a genotoxic carcinogen is risk-free, and to estimate risks from low exposures by extrapolating down linearly from doses at which carcinogenic responses were observed in animal studies (Albert, 1994). It is not surprising that, as the need arose for conducting risk assessments on chemical carcinogens, the approach taken was to borrow heavily from what was known about radiation carcinogenesis. Indeed, today a cornerstone of assessing the carcinogenic risks of chemicals is that if the chemical or its metabolites cause gene mutations by interacting directly with DNA, they cause cancer in a manner analogous to radiation, and the dose-response for the chemical can be assumed to have a linear, no-threshold relationship.

To take account of these considerations, the USEPA modified its risk assessment methodology for carcinogens. If the mode of action (MOA) of a substance is sufficiently well understood and indicates that the carcinogenic dose response is threshold or nonlinear at low doses, the risk assessment is handled in the same way as for a non-carcinogen. Otherwise, quantitative estimates of low-dose risk are calculated by linearly extrapolating downward from a point of departure (PoD) (USEPA, 2005). A similar approach is used in the Netherlands (VROM, 1989), Norway and Germany. Other countries use no quantitative extrapolation but implement the ALARA (As Low As Reasonably Achievable) principle (Neumann, 2009).

The classical standard animal study protocol for carcinogenesis testing is a lifetime 2-year rat or 18-month mouse study using continuous exposure to three dose groups and an untreated control. The highest dose tested is the maximum tolerated dose (MTD), with usually half and a quarter of this dose as the additional dose levels tested. The USEPA proposes that the dose that causes a 10% increase in tumour incidence (ED10) or its lower confidence limit (LED10) is used as the PoD for extrapolating downward (using a simple linear non-threshold model as a default or a non-linear model if this is supported by mechanistic information). In the Netherlands, the lowest dose resulting in tumours in the animal study is linearly extrapolated to zero exposure to derive a dose related to a one-in-a-million chance of acquiring a tumour (acceptable or tolerable risk level), which is used as the basis for human risk management. The European legislation for REACH uses either an interpolated dose equal to 25% increase in tumour incidence (T25) or the BMDL10, i.e. the lowest limit of the confidence interval of the benchmark dose associated with a 10% increase in tumour incidence (ECHA, 2012).

This process has been heavily disputed since its introduction in the 1980s. First, the mathematical extrapolation (through a simple linear model or a more complex model) from very high MTD-related doses (experimental/observed range) to the very low dose range associated with theoretical risk levels of $1{:}10^6$ or $1{:}10^5$ has a high level of uncertainty, because it involves extrapolation up to 5-6 orders of magnitude below the experimental range. This also results in differing risk estimates for the same

substance, depending on the model chosen. Second, the shape of the dose-response below the experimental range is unknown. Third, the assumed low-dose linearity in the dose-response appears biologically implausible from our knowledge of basic biology and physiology. Fourth, the significance of tumours induced in animals at or around the MTD (i.e. in the presence of significant generalised toxicity) to humans exposed to much lower doses is uncertain. Fifth, the risk estimates so derived provide a false sense of accuracy. Moreover, the establishment of acceptable or tolerable risk levels (e.g. 1 in 10^6 versus 1 in 10^5) is beyond purely scientific considerations and requires that due account is taken of policy and societal considerations.

This two-track approach to chemical risk assessment, in which quantitative estimates are provided only for genotoxic carcinogens is a well-established regulatory practice. The decision about which track is appropriate for a given toxicant turns on whether or not a threshold (or a nonlinerar dose-response) exists, which is often a very contentious issue.

It should be noted that the linear low dose extrapolation approach adopted for genotoxic carcinogens was developed at a time when modern insights into mechanisms of tumour initiation, promotion and progression and of physiological defence mechanisms were yet to be revealed. First, the body has a wealth of absorption, distribution, metabolism and excretion (ADME) mechanisms in place to detoxify and remove xenobiotic compounds, which reduces the chance of a genotoxic molecule to reach the DNA. Alternatively, metabolic conversion of inactive compounds to toxic derivatives may occur, which requires metabolic enzyme induction, which will only occur above a threshold of exposure. Secondly, if DNA damage is inflicted, various DNA repair mechanisms are in place to undo the damage, protecting the cell from acquiring DNA mutations. Thirdly, the carcinogenic process is now known to consist of a cascade of cancer-promoting changes, which all need to occur before cancer arises. The likelihood that all of these changes occur in concert without being repaired by homeostatic mechanisms is very low, thereby further reducing the chance that exposure to a single genotoxic molecule will lead to cancer, and implying that a biological threshold must exist. Overall, therefore, there is a growing amount of evidence for the existence of thresholds of adversity even for directly-acting genotoxic agents, which challenges the scientific validity of the linear extrapolation approach to the risk assessment of genotoxic carcinogens (Pratt et al., 2009).

The EFSA Scientific Committee has expressed serious reservations about linear high to low dose extrapolation far beyond the tested range for human low-dose risk estimation for genotoxic carcinogens (EFSA, 2005). The Scientific Committee recommends the use of a different approach, known as the margin of exposure (MoE) approach. The MoE provides a simple and practical approach that avoids the scientific uncertainties associated with the selection of a mathematical model for low dose extrapolation, and doubts about its biological relevance at low doses. The MoE approach uses a reference point, often taken from an animal study (but also from human data) corresponding to a dose that causes a low but measurable tumourigenic response. This reference point is then compared with exposure estimates in humans. As the reference point, the Scientific Committee recommends the use of the BMDL10 (benchmark dose lower confidence limit 10%) which is an estimate of the lowest dose which is 95% certain to cause no more than a 10% cancer incidence in rodents. The Scientific Committee gives also guidance on how to interpret the MoE. The following aspects are proposed to be taken into account when interpreting the numerical value of the MoE: interspecies differences, intraspecies differences, the nature of the carcinogenic process (particularly human variability in cell cycle control and DNA repair) and the significance of the reference point (not considered as a surrogate for a threshold). The Scientific Committee is of the view that in general an MoE of 10,000 or higher, if it is based on the BMDL10 from an animal study, would be of low concern from a public health point of view and might be considered as a low priority for risk management action. The value of 10,000 is obtained by multiplying the default 100-fold factor for interspecies and intraspecies difference with an additional default 100-fold factor for uncertainties in the nature of the carcinogenic process and significance of the reference point.

It is now well accepted that the presence or absence of a mathematical or absolute threshold cannot be proven or disproven experimentally as this would require an infinitely sensitive method with an infinitely large number of animals and an infinitely small dose, down to one molecule (Slob, 1999; Crump, 2011; Rhomberg *et al*., 2011). Science is not capable of determining the shape of the dose response at very low doses. Hypotheses regarding the existence or nonexistence of thresholds are beyond the ability of science to resolve. Continuing to expend energy and time debating the irresolvable issue of thresholds is detrimental to a logical and workable, comprehensive approach to risk assessment.

However, despite these assertions, the debate over the nature of the exposure (dose)-response relationship has now been extended from cancer to a wide range of non-cancer endpoints (White et al., 2009; NRC, 2009), including endocrine disruption (Blair *et al*., 2001; Zoeller *et al*., 2012). It is debated whether agents causing non-cancer toxicity at high exposure levels should, as a default, be presumed to cause some degree of risk of these same endpoints at any dose, no matter how low. The basis for assuming that all exposure-response relationships are linear and non-thresholded include (1) the general "additivity-to-background" argument, which assumes that if an agent enhances an already existing disease-causing process, then even small increases in exposure concentration and/or duration increase disease incidence in a linear manner; and (2) the "infinite sensitivity of the population" argument, which assumes that there would always be at least one very sensitive individual in the population which will show an adverse response even to one molecule of a chemical agent.

In response to these views, Rhomberg *et al.* (2011) argue that the no-threshold proposal for non-cancer toxicity is at odds with decades of experience and repeatable observations in exposure-response relationships in pharmacology and toxicology and with the basic tenets of homeostasis. The presence of homeostatic and defence mechanisms, and the redundancy of cellular targets mean that a minimum degree of interaction of the chemical agent with the critical sites must be reached in order to elicit a toxicologically relevant effect. Below this critical level of interaction (threshold of adversity), homeostatic mechanisms would be able to counteract any perturbation produced by xenobiotic exposure, and no structural or functional changes would arise. It is also disputed that the infinite sensitivity of the population argument is an abstract mathematical concept, which has no corroboration from real world observations. They conclude that human risks at low doses, if they exist, are too rare to observe directly, and so inferences must be made that depend on their validity on invoking wider biological understanding of what should be expected to occur at low levels of human exposure. They also conclude that biology predicts that thresholds of adversity exist and are the rule, rather than the exception, for all endpoints.

On the basis of these considerations, Crump (2011) proposes a harmonised PoD/SF approach to risk assessment that could be applied in all cases. The approach includes a "risk-reduction" factor determined from MOA information which would account for how far below the PoD a dose must be in order to be reasonably safe, and a "severity adjustment" factor to ensure that more serious toxic effects would be regulated more stringently than less serious ones, and vice versa.

Alternatively, the MoE approach proposed by the EFSA Scientific Committee for the risk assessment of genotoxic carcinogens (EFSA, 2005) could also be used as a harmonised risk assessment methodology applicable to all substances, effects or modes of action.

There are many examples of situations where, because of too many gaps in the toxicological dataset to allow the establishment of a robust reference value, the MoE approach was used. The MoE is calculated as the ratio between a defined point on the dose-response curve for the adverse effect, often the NOAEL or BMD or BMDL, and the human exposure estimate.

In evaluating whether an MoE is high enough to conclude that a chemical is unlikely to cause harm, the usual toxicokinetic and toxicodynamic uncertainties related to species differences and potential human variability associated with the data must be taken into account. The reliability of the MoE determined for a substance would depend upon the adequacy of the database. Choosing the acceptable margin (minimal MoE) between the critical point on the dose-response curve and the exposure presents similar difficulties to choosing a SF in an ADI calculation. However, compared with a reference value, the MoE also takes also into account the uncertainties in the exposure estimate.

In calculating the MoE, there is no necessity to use the NOAEL, and any appropriate defined point on the dose-response curve would be suitable. The MoE can be used without making any implicit assumptions about safety.

It is often said that the MoE that would be considered acceptable is a societal judgement and should not be determined by risk assessors alone. Indeed, use of the MoE can enable complete separation of risk assessment from risk management. The risk assessor would provide the risk manager with the magnitude of the MoE and the risk manager would decide on its acceptability. However, risk assessors have the responsibility to inform risk managers on the nature of the critical hazard, the quality of the toxicological data and the uncertainties inherent in the data used for the exposure estimates.

7.5. Consideration of uncertainty

Introduction

The consideration of uncertainty in setting health-based guidance values is a recommended part of several risk assessment schemes. It is a particularly important aspect for communicating with risk managers and members of the public (ref COT report 2007)). Although it might be impossible to identify or quantify all the uncertainties in an assessment it is helpful if some indication of the types and magnitudes of the uncertainties can be provided. However, in reviewing recent texts (EFSA reports, USEPA REDs post 2006) emerging from such schemes it was clear that, in general, uncertainty was mentioned specifically only when there was a high level of uncertainty such as the absence of a critical study or there was some unusual aspect within a particular study (e.g. dose spacing). One exception was the JMPR where there would be text to indicate if an ARfD was considered to be highly conservative (e.g. due to the use of a repeat dose study) and could probably be refined by further data from a study of more relevant duration. The USEPA regularly commented on uncertainties regarding the exposure aspects of the human health risk assessment (where the available data were more amenable to being plotted as distributions and specific centile values identified) but not normally on the toxicology, other than for missing studies or in respect of the FQPA considerations.

The consideration of uncertainty has been reviewed by a number of bodies recently. The EFSA Scientific Panel produced a report in 2006[39] relating to dietary exposure estimates. This was supplemented by the EFSA PPR panel advice which included a discussion of uncertainty in a report on MRLs (EFSA, 2008[40]), and a consideration on transparency that considered that uncertainty in the derivation of the toxicology end-points should also be described (EFSA, 2009[41]). A specific example of an approach to describe uncertainties associated with toxicological end-points is given in the EFSA PPR panel report on cumulative effects of exposures to triazole pesticides (EFSA 2009).

[39] http://www.efsa.europa.eu/en/efsajournal/doc/438.pdf
[40] http://www.efsa.europa.eu/en/efsajournal/pub/438.htm
[41] http://www.efsa.europa.eu/en/rapractice/ratransparency.htm

Presentation of uncertainty

The level of uncertainty can vary with different parts of the assessment and it can be difficult to combine all these uncertainties into an overall consideration. By presenting uncertainties separately it is possible to identify the critical aspects that should be addressed to lead to a refinement of the overall assessment. The presentation of uncertainty in the derivation of toxicological end-points can be addressed in several ways.

- Qualitatively : 'Because the magnitude of the effect at the LOAEL is small and there is a large gap to the NOAEL it is likely that the NOAEL will be conservative'. 'This compound is probably a human mutagen as there are positive results *in vitro* and in animal studies and there is no reason to assume humans will respond differently'.

- Semi-quantitatively: The use of one or more plus (+) or minus (-) signs to indicate the likelihood of a value being an under-estimate or an over-estimate of the true value. For example -/+++ would indicate there is a greater chance of the value being an over-estimate than an under-estimate; whereas --/++ would indicate the value was as likely to be an under-estimate as an over-estimate and that the magnitude of the 'uncertainty' was moderate. This is the approach outlined in the EFSA triazole opinion (Table 44 of EFSA, 2009).

- Quantitatively: The use of values from BMD analyses and relating the value chosen for the end-point to the range of the BMDL to BMDU. 'The NOAEL is 1.5 mg/kg bw/day, the BMDL is 1.6 mg/kg bw/day and the BMDU is 7.4; therefore there is a high level of confidence that there will be no adverse effects at the NOAEL.'

- A combination of approaches: Specific analyses are performed and the probability of a particular outcome is converted into an agreed phrase that covers the quantitative and qualitative considerations: 'It is likely that the threshold for this effect is above this value' as the probability value was between 60 and 85%.

  In many cases one approach will not address all aspects as some questions will be addressed qualitatively, e.g. is this compound neurotoxic (yes or no?); some quantitative, e.g. what is the threshold dose for neurotoxicity in rats; some a mixture, e.g. what is the likelihood that an exposure of xx will be neurotoxic in humans.

A report for the Food Standards Agency in the UK (ref TA1056[42]) and subsequent discussions at a workshop and a Meeting of the Committee on Toxicity[43] have considered the presentation of uncertainty to the general public and involved discussions with social scientists as to the best approach. There was no definitive conclusion but the outcomes included the use of standardised phrases combined with some degree of numerical description and a framework was proposed that addressed both qualitative and quanitiative approaches; however, the framework needs to be tested with worked examples[44]. The use of standardised phrases linked to numerical values is also used by the UK Pesticide Residues in Food Committee when providing an assessment of the risks of adverse effects when pesticide residues found during surveillance exercises result in predicted intakes above the ARfD [45]. There is also evidence that the methods of presenting the uncertainties should be tailored to the audience i.e. the approach that suits risk managers might not be relevant to explaining to members of the public.

[42] http://www.food.gov.uk/science/research/foodcomponentsresearch/riskassessment/t01programme/t01projlist/t01056/
[43] http://cot.food.gov.uk/pdfs/cotmins22jun2010.pdf
[44] http://cot.food.gov.uk/pdfs/tox201019.pdf
[45] http://www.pesticides.gov.uk/Resources/CRD/PRiF/Documents/Results%20and%20Reports/2011/PRiFAnnualReport2011WEBFINAL.pdf

Why is uncertainty not addressed in pesticide evaluations?

It is clear that there is a sound scientific and presentational basis for providing some indication of the uncertainties surrounding the toxicological evaluation of chemicals. However, currently this is not done for routine pesticide evaluations from any agency. Reasons for this could include:

- the template documents do not include a section for uncertainty;
- there are no agreed methods or guidance;
- expressing uncertainty would be of no immediate value to the regulatory process. Risk managers do not ask for an indication of uncertainty. In some instances, e.g. in the EU, the legislative framework (pesticide uniform principles) precludes a consideration of uncertainty in that if the exposure exceeds the reference dose then authorisation is not permitted even if the exceedance is small (e.g. 5%) and all indications are that the assessment includes many conservative aspects.

Proposals for addressing the description of uncertainty.

- Although there is not a single 'best' approach to presenting the uncertainty around toxicological assessments, it would not be difficult to use currently available techniques to at least give some indication of what the uncertainties are and their magnitude / direction.
- Templates for preparing summary documents (OECD, EU, JMPR refs) should include a section for uncertainty with a pro-forma of standard questions (e.g. completeness of the database; consistency of effects across species, particularly sensitive species and the impact on uncertainty factors; magnitude of effect at critical LOAEL and dose spacing to NOAEL; [range for BMDL – BMDU]) to encourage some indication of where uncertainties lie.
- Further work to determine the best way to present uncertainties to risk managers and the general public is required. Risk managers should give guidance on what they need in terms of an uncertainty assessment and how they would like it to be presented.

8. OVERALL CONCLUSIONS AND RECOMMENDATIONS

8.1. Conclusions

Analysis of current approaches for the derivation of reference values

- The current procedures (i.e the application of a safety factor to a NOAEL) for deriving health based guidance values (ADI, AOEL, ARfD) are widely accepted and applied around the world. There is no strong evidence that they are not providing protection to the human population exposed to the classes of chemicals covered by these values. Alternative approaches such as Margin of Exposure (MoE) might have application in specific cases (e.g. genotoxic carcinogens) but do not have as broad an applicability as the setting of a formal health-based guidance value. Where differences in values for particular chemicals exist between evaluating agencies, these can be attributed primarily to one or more of: subjective factors (e.g. expert judgement used to interpret the data); legislative restrictions (e.g. the use of human data; minimum safety factors to be applied); or variations in the supporting database.

- Different risk management considerations mean that there will never be complete consistency across agencies worldwide. However, to improve consistency within a particular scheme there should be clear guidance on what approaches are preferred and which ones are not applicable.

- A comparison of JMPR and EFSA ADI values shows reasonable consistency (34/57 values were common other than differences in the use of rounding) and no clear pattern behind the differences. However, for ARfD setting, while 24/57 values were common, JMPR set higher values than EFSA for 23 of the 33 compounds and concluded that an ARfD was not necessary for a further 7 compound. This difference in approach can, and has, caused difficulties in the assessment of consumer intakes of certain pesticide residues and the setting of MRLs.

- Current guidance used by EU evaluators for deriving health-based guidance values results in approximately 50% of the initial proposals made in the DAR being confirmed after peer review. This indicates that the available guidance is valuable but could be improved in certain areas. Any changes to the guidance should ensure they do not compromise current procedures where they work.

- The default 100 fold SF was used in the vast majority of guidance value derivations.

- Tighter dose spacing at the low end of the dose response could significantly reduce the uncertainty surrounding the NOAEL. For a significant number of compounds the margin between NOAEL and LOAEL was >5.

- Analysis of the derivation of ADIs identified a number of aspects that were regularly critical in reaching conclusions:
    - the interpretation of adversity for increased liver weight and hypertrophy; and for reduced body weight;
    - the derivation of combined NOAELs;
    - the use of extra safety factors for use of a LOAEL or severity of effects. The basis for the extra factor was normally not substantiated.

- Analysis of the derivation of AOELs identified a number of aspects that were regularly critical in reaching conclusions:
    - the interpretation of adversity for increased liver weight and hypertrophy, and for reduced body weight;
    - the derivation of combined NOAELs;
    - the use of extra safety factors for use of a LOAEL or severity of effects. The basis for the extra factor was normally not substantiated;
    - determination of the appropriate study duration for the use pattern of the pesticide;
    - concluding on the appropriate oral absorption value, particularly when this was near the 80% cut-off and when there was a significant biliary excretion but the liver was not the target organ. This is important as an oral absorption correction was used in approximately one third of AOELs.

- Analysis of the derivation of ARfDs identified a number of aspects that were regularly critical in reaching conclusions:
    - the interpretation of adversity for reduced body weight;
    - the derivation of combined NOAELs;
    - the use of extra safety factors for use of a LOAEL or severity of effects. The basis for the extra factor was normally not substantiated.

- A number of end-points such as cardiovascular or immunotoxic effects were hardly ever used in determining reference values, but it is unclear if this is because specific investigations for these effects did not form part of the study protocols or they are not sensitive end-points. The latter is unlikely as analyses of study NOAELs by critical effect indicated no specific organ or tissue was clearly more or less sensitive than others. Immunotoxicity investigations are included in current test protocols but specific in-life measures of the cardiovascular system parameters are not included.

AAOEL derivation

- Analysis of data on exposures relevant to an AAOEL indicated that under most circumstances the dermal route would be the primary route of exposure to consider. A comparison of results from acute dermal and acute oral toxicity studies and dermal developmental toxicity studies showed that while single dose effects were less prevalent following dermal exposure, in some instances effects were produced that were equivalent to those seen with oral dosing. Therefore the default position should be that an AAOEL assessment is applied in all cases, certainly, if an ARfD has been set. Certain end-points such as vomiting and diarrhoea used to set ARfDs could be set aside if shown to be secondary to local effects, but in such cases other potentially relevant acute endpoints would need to be considered. In the event of there being extensive inhalatory exposure of spray particles, this would be addressed by the ARfD as, due to particle sizes of typical spray nozzles, the internal dose would be from ingestion following mucociliary clearance. These indicate that a systemic AAOEL derived by correcting an ARfD for oral absorption would be an adequate first tier approach with potential for refinement if appropriate data are available.

  If an ARfD was considered and determined to be not necessary there is no reason why an AAOEL would be required.

New/alternative approaches

- Improvements to the procedures for deriving the guidance values should focus on the use of new approaches to refine the derivation of health based guidance values and the improvement of guidance on aspects that are used routinely and / or regularly result in discussions during peer review:
    - use of the benchmark dose approach to provide an indication of the uncertainties surrounding the PoD, and to providing an indication of the numerical value of additional factors when using a LOAEL;
    - potential to refine the reference values by the use of chemical specific adjustment factors or allometric scaling factors;
    - providing additional guidance to improve the consistency and scientific relevance of the values, especially these are based on effects of uncertain toxicological significance or human relevance such as liver hypertrophy; body weight deficits and species-specific findings; and additional guidance on the application of safety factors >100.
    - developing a consistent approach to the derivation of combined NOAELs for two or more studies.
    - provide a template to help in the presentation and identification of uncertainties.

BMD Analysis

- BMD modelling provides a superior scientific approach compared with the NOAEL for the derivation of a PoD for a reference dose, as it makes use of all the dose-response data and provides an indication of the uncertainty (confidence) hidden in the PoD. BMD modelling is not at present a tool for routine use by human health risk assessors. The software programmes are still being developed, can be unreliable and they require a degree of specialist knowledge for reliable interpretation. However, despite these practical problems, the scientific supremacy of the BMD approach compared to the NOAEL method should be an incentive to apply it at least as a higher-tier or supplementary method when the critical study for the derivation of a reference value has been identified. The application of the BMD approach at this stage of the risk assessment process will generate a more robust and more transparent risk estimate with an indication of the associated uncertainty, which, in turn, could lead to improved risk communication between risk assessors, risk managers, policy-makers and the public.

Allometric scaling

- Allometric scaling, as opposed to the application of a default 10 fold factor to extrapolate from all animal species to average humans, is used in the derivation of DNELs under REACH and the USEPA has also moved to this approach for pesticide evaluations. The approach has theoretical shortcomings when applied to compounds with a high first pass metabolism or reactive metabolites. However, it does provide a basis for correcting between animal species with differing metabolic requirements and could be used as a second-tier approach, especially if dog data are the basis for the establishment of a reference value. Allometric scalling should enable more refined and scientifically-based risk assessments: the historical default interspecies factor of 10 is arbitrary, whereas species-specific allometric scaling factors are supported by experimental observations and by fundamental physiological and biochemical processes that underlie the kinetics of toxicants.

CSAFs

- The use of chemical-specific assessment factors has been used by some organisations (e.g. JMPR) in deriving reference values. It requires additional data on relative toxicokinetics and dynamics across species and within human sub-populations. Because of the requirement for additional information it cannot be applied generically.

PBPK Modelling

- Other approaches to addressing the uncertainty in moving from a NOAEL or BMD to a reference value, such as PBPK /PBTK modelling, can be used case-by-case to refine a standard assessment but are too data intensive to be used routinely.

Adverse or adaptive nature of effects observed in standard toxicity studies

- There are no criteria that can be used to provide clear guidance on whether particular effects seen in toxicity studies are adaptive, adverse or not relevant to humans. For liver weight increases there is a general acceptance that an increase of <10% is not adverse. For body weight reductions, decreases < 10% are also not considered adverse and for inhibition of AChE, reductions of < 20% are not considered adverse. A number of other effects (e.g. to the kidney or thyroid) that have some mechanisms of toxicity indicating that might be not relevant to humans need to be considered using a weight of evidence approach. However, until demonstrated otherwise, the default assumption is that effects seen in animals are relevant to humans.

8.2. Recommendations

- BMD analyses should be performed on the critical studies and endpoint(s) used to set health-based guidance values. As a minimum, BMD analysis should be used to provide an indication of the uncertainty surrounding the NOAEL. Such analyses should always be performed if a LOAEL is used, to provide an indication of the magnitude of any extra safety factor to be applied. BMD software programmes are available and undergoing constant development, but adequate expertise in performing BMD analyses is not currently available in all regulatory agencies. Therefore, in the near future a central agency should take the leading role in promoting and harmonising BMD software programmes and their use and in providing training to national authorities.

- The OECD (and equivalent bodies / organisations) should consider revisions to toxicity study test guidelines to determine if some of the basic requirements that have remained unchanged for decades need to be updated. Such considerations should include measurements of basic cardiovascular system parameters and the use of more dose groups without increasing the total number of animals; this would provide more information on the dose-response relationship and reduce uncertainties when BMD analyses are performed.

- There is good evidence to support the contention that certain findings in laboratory animals are either adaptive or not relevant to humans. However, there are often a number of alternative modes / mechanisms of action for a single finding and not all of them can be dismissed as being adaptive or without human relevance. It is therefore the responsibility of the submitting company to present a case, specific to the molecule and effect, demonstrating the adaptive nature or lack of human relevance. Such cases should be presented in a structured way (e.g. IPCS MoA and human relevancy framework by Boobis *et al*., 2008).

- Some indication of the uncertainty surrounding a health-based guidance value should always be given, particularly if there is clear scope for refinement by the generation of additional information. It is unclear what form this should take, as a form that is of use to risk managers might be of limited relevance to members of the public; hence, further work on this aspect should be performed.

- EU guidance on setting ADIs should be developed. The basis for this already exists in documents such as Environmental Health Criteria 240 and the guidance for the setting of AOEL and ARfD. It should include advice on criteria required for discriminating between adverse effects versus non-adverse / non-human-relevant for commonly used endpoints. It should indicate when alternative / new approaches, such as BMD or CSAF, can be used either routinely or to refine values derived by standard approaches. The basic principles for refining ARfDs have been described by Solecki et al (2005) and could be applied to refining ADIs.

- The EU AOEL guidance should be updated to provide specific guidance on areas that cause most discussion e.g. appropriate duration of study, which needs to be linked with information on exposure patterns; and oral absorption particularly whether actual values rather than the ≥ 80% approach should be used.

- EU guidance on ARfDs should be updated to include aspects related to refinement (as described in Solecki *et al.*). It should also clearly define criteria for determining if an ARfD is not required. The guidance should be clear on the use, or not, of human data and the application of CSAFs, as these are the key differences between the current EFSA approach and that used by JMPR.

- EU guidance on deriving AAOELs should be developed.
    - The default assumption should be that an AAOEL should be considered for all active substances.
    - For active substances where an ARfD has been set this can be the initial basis for the AAOEL, subject to correction for oral absorption.
    - If an ARfD is based on gastrointestinal effects or reduced food consumption these should be considered as applicable to the setting of an AAOEL until a case has been presented to show that they are not relevant to non-oral exposures. If this is the case, alternative acute end-points need to be considered.
    - If an ARfD has been deemed unnecessary, the default assumption is that the same conclusion can apply to the need for an AAOEL.
    - If an active substance was reviewed before ARfDs were routinely considered and an acute non-dietary assessment is required, an AOEL can be used as an initial value pending a re-evaluation of the database. Moving from a 75%ile exposure estimate to a 95%ile exposure value (as normally required for acute effects) will typically result in a six-fold increase. Therefore it is likely that a number of acute assessments based on an AOEL will require refinement.
    - Before any additional animal studies are performed to refine an AAOEL, exposure scenarios and dermal absorption should be refined as far as possible.
    - If an AAOEL (which is still a systemic reference value) is not required, then there should still be a consideration of the potential of local effects to workers, bystanders

and residents for products or dilutions classified (under the CLP Regulation) for irritation/corrosivity and/or sensitisation.

REFERENCES

Albert R (1994). Carcinogen risk assessment in the U. S. Environmental Protection Agency. *Crit Rev Toxicol* 24:75–85.

Alexander, J, Benford, D, Boobis, A, Eskola, M, Fink-Gremmels, J, Fürst, P, Heppner, C, Schlatter, J and van Leeuwen, R (2012). Risk assessment of contaminants in food and feed. *EFSA Journal* 10(10): s1004.

Andrew, D (2005). PSD guidance document: interpretation of liver enlargement in regulatory toxicity studies. York, UK.
http://www.pesticides.gov.uk/Resources/CRD/Migrated-Resources/Documents/A/ACP_Paper_on_the_interpretation_of_Liver_Enlargement.pdf

Boobis, AR, Cohen, SM, Dellarco, VL, McGregor, DB, Meek, ME, Vickers, C, Willcocks, D and Farland, W (2006). IPCS framework for analyzing the relevance of a cancer mode of action for humans. *Crit.Rev.Toxicol.* 36:781-792.

Boobis, AR, Doe, EJ, Heinrich-Hirsch, B, Meek, EM, Munn, S, Ruchirawat, M, Schlatter, J, Seed, J and Vickers, C. (2008). IPCS framework for analysing the evincence of a non-cancer mode of action for humans. *Crit.Rev.Toxicol.* 38: 87-96.

Barton, HA, Chiu, WA, Setzer, RW, Andersen, ME, et al. (2007). Characterizing uncertainty and variability in physiologically-based pharmacokinetic (PBPK) models: state of the science and needs for research and implementation, Toxicol Sci, (in press). [Toxicol Sci, Advance Access published May 4, 2007 doi: 10.1093/toxsci/kfm100].

Bharadwaj, LA (2009). Cardiac Toxicity (Chapter 57) in *General and Applied Toxicology 3rd edition/ editors Ballantyne B; Marrs TC & Syversen T.* John Wiley and Sons, Chichester, UK ISBN 978-0-470-72327-2. pp 1329 – 1350

Blair RM, Fang H, Gaylor D and Sheehan DM (2001). Threshold analysis of selected dose-response data for endocrine active chemicals. *APMIS* 109:198-208.

Boobis, AR; Doe, JE; Heinrich-Hirsch, B *et al* (2008). IPCS framework for analysing the relevance of a noncancer mode of action in humans. *Crit Rev Toxicolol* 38(2): 87 – 96.

Brown, KG. and Erdreich, LS (1989). Statistical uncertainty in the no-observed adverse effect level. *Fund Appl Toxicol*, 13: 235–244.

Brown, RP, Delp, MD, Lindstedt, SL, Rhomberg, LR, and Beliles, RP. (1997). Physiological parameter values for physiologically based pharmacokinetic models. *Toxicology and Industrial Health* 13: 407-484.

Clewell III, HJ, and Andersen, ME (1987). Dose, species and route extrapolation using physiologically-based pharmacokinetic modeling. *Drinking Water and Health* 8: 159-184

Crump, KS. (1984). A new method for determining allowable daily intakes. *Fundam. Appl. Toxicol.* 4:854-71.

Crump KS (2011). Use of threshold and mode of action in risk assessment. *Crit Rev Toxicol* 41(8):637–650.

Dedrick, RL (1973). Animal scale-up. *J Pharmacokinet Biopharm*. 1(5):435-61.

Dedrick, RO, Bischoff, KB, and Zaharko, DS (1970). Interspecies correlation of plasma concentration history of methotrexate (NSC-740). *Cancer Chemother Rep Pt*. 54: 95-101.

Dedrick, RL, and Bischoff, KB (1980). Species similarities in pharmacokinetics. *Fed Proc* 39: 54-9.

Dellarco, VL, McGregor, D, Berry, C, Cohen, SM and Boobis, AR (2006). Thiazopyr and thyroid disruption: case study within the context of the 2006 IPCS human relevance framework for analysis of a cancer mode of action. *Crit.Rev.Toxicol.* 36:793-801.

Dorne, J.L.C.M, 2010. Metabolism, variability and risk assessment. Toxicology, 268, 156-164.

Dorne J.L.C.M, Walton K, Renwick A.G, 2004. Human variability in xenobiotic metabolism and pathway-related uncertainty factors for chemical risk assessment: a review. Food and Chemical Toxicology 43, 203–216.

EC (European Commission), 2006. Draft Guidance for the Setting and Application of Acceptable Operator Exposure Levels (AOELs) (SANCO 7531 - rev.10, 2006).

EFSA SC (2005). Opinion of the Scientific Committee on a request from EFSA related to A Harmonised Approach for Risk Assessment of Substances Which are both Genotoxic and Carcinogenic. *The EFSA Journal,* 282 :1-31.

EFSA Panel on Contaminants in the Food Chain (CONTAM) (2010). Scientific Opinion on Lead in Food. *EFSA Journal* 8(4):1570. [147 pp.].

EFSA Panel on Contaminants in the Food Chain (CONTAM) (2011a). Scientific Opinion on tolerable weekly intake for cadmium. *EFSA Journal* 9(2): 1975. [19 pp.].

EFSA Panel on Contaminants in the Food Chain (CONTAM) (2011b). Scientific Opinion on the risks for animal and public health related to the presence of *Alternaria* toxins in feed and food. *EFSA Journal* 9(10): 2407. [97 pp.].

EFSA Panel on Contaminants in the Food Chain (CONTAM) (2011c). Scientific Opinion on the risks for public health related to the presence of T-2 and HT-2 toxin in food and feed. *EFSA Journal* 9(12): 2481. [187 pp.].

EFSA (European Food Safety Authority), 2009.
Guidance of the Scientific Committee on a request from EFSA on the use of the benchmark dose approach in risk assessment. The EFSA Journal (2009) 1150, 1-72

EFSA (2009a). Scientific opinion of the Panel on Contaminants in the Food Chain on a request from the European Commission on cadmium in food. *EFSA Journal* 980: 1-139.

EFSA Scientific Opinion (2009b). Use of the benchmark dose approach in risk assessment. Guidance of the Scientific Committee (Question No EFSA-Q-2005-232) Adopted on 26 May 2009

EFSA Panel on Plant Protection Products and their Residues (PPR); 2010. Scientific Opinion on Preparation of a Guidance Document on Pesticide Exposure Assessment for Workers, Operators, Bystanders and Residents..2010;8(2):1501, 24 pp.

EFSA (2012a) Scientific Opinion on exploring options for providing advice about possible human risks based on the concept of the threshold of toxicological concern. EFSA journal 10 (7): 2750 http://www.efsa.europa.eu/en/efsajournal/pub/2750.htm

EFSA (2012b) Scientific Opinion on evaluation of the toxicological relevance of pesticide metabolites for dietary risk assessment. EFSA journal 10 (7): 2799 http://www.efsa.europa.eu/en/efsajournal/doc/2799.pdf

European Chemicals Agency (ECHA). (2008). Guidance for the implementation of REACH. Guidance on information requirements and chemical safety assessment Chapter R.8: Characterisation of dose [concentration]-response for human health. Available at: http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_en.htm#A

European Chemicals Agency (ECHA) (2012). Guidance on information requirements and chemical safety assessment. Chapter R.8: characterisation of dose [concentration]-response for human health. European Chemicals Agency, Helsinki. http://echa.europa.eu/documents/10162/13632/information_requirements_r8_en.pdf

European Commission Groups of Specialised Experts in the fields of Carcinogenicity, Mutagenicity and Reprotoxicity. (1999). Summary record: non-genotoxic thyroid carcinogens in the rodent bioassay. ECBI/49/99, 1-2 September 1999.

FAO/WHO (2004) Pesticide residues in food. Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues. Rome, Food and Agriculture Organization of the United Nations (FAO Plant Production and Protection Paper, No. 178. http://www.fao.org/ag/AGP/AGPP/Pesticid/JMPR/DOWNLOAD/2004_rep/report2004jmpr.pdf.

FAO/WHO (2009). Environmental Health Criteria 240. Principles and methods for the risk assessment of chemicals in food. World Health Organization, Geneva. http://www.who.int/foodsafety/chem/principles/en/index1.html

FAO/WHO (2011). Safety evaluation of certain food additives and contaminants / prepared by the seventy-third meeting of the Joint FAO/WHO Expert Committee on Food Additives (JECFA). WHO Food Additives Series No. 64 pp 305-380. World Health Organization, Geneva. http://whqlibdoc.who.int/publications/2011/9789241660648_eng.pdf

Galli C.L, Marinovicha M, Lotti M, 2008. Is the acceptable daily intake as presently used an axiom or a dogma? Toxicology Letters, 180, 93–99.

Gargas, ML, Burgess, RJ, Voisard, DE, Cason, GH, and Andersen, ME (1989). Partition coefficients of low molecular weight volatile chemicals in various liquids and tissues. *Toxicology and Applied Pharmacology* 98: 87-99.

Gaylor, DW. (1988). Applicability of cancer risk assessment techniques to other toxic effects. *Toxicol. Ind. Health*. 4(4): 453-9.

Gaylor, DW (1992). Incidence of developmental defects at the no observed adverse effect level (NOAEL). *Regul Toxicol Pharmacol*, 15: 151–60.

Hall, AP; Elcombe, CR; Foster, JR et al (2012) Liver hypertrophy: A review of adaptive (adverse and non-adverse) changes – Conclusions from the 3rd International ESTP expert workshop. *Toxicologic pathology* 40(7): 971 – 974

Hard, G and Khan, K (2004). Invited review: a contemporary overview of chronic progressive nephropathy in the laboratory rat, and its significance for human risk assessment. *Toxicologic Pathology*, 32, 171-180.

Hard, GC, Whysner, J, English, JC, Zang, E and Williams, GM (1997). Relationship of hydroquinone-associated rat renal tumors with spontaneous chronic progressive nephropathy. *Toxicologic Pathology*, 25: 132-143.

Health and Safety Executive (2011). EH40/2005 Workplace Exposure Limits containing the list of workplace exposure limits for use with the Control of Substances Hazardous to Health Regulations (as amended), Second Edition. Published by HSE.
http://www.hse.gov.uk/pubns/books/eh40.htm.

Hill, RN, Erdreich, LS, Paynter, OE, Roberts, PA, Rosenthal, SA. and Wilkinson, C. F. (1989) Thyroid follicular cell carcinogenesis. *Fund.Appl.Toxicol.* 12: 629-697,

Hill, RN, Crisp, TM, Hurley, PM, Rosenthal, SL. and Singh DV (1998). Risk assessment of thyroid follicular cell tumours. *Environ.Health Perspect*. 106: 447-457.

Hinton, RH; Grasso, P & Marrs TC (2009). Hepatotoxicity (Chapter 59) in *General and Applied Toxicology 3rd edition/ editors Ballantyne B; Marrs TC & Syversen T.* John Wiley and Sons, Chichester, UK ISBN 978-0-470-72327-2. pp 1369 – 1410.

IARC (1999). Consensus Report in *Species Differences in Thyroid, Kidney and Urinary Bladder Carcinogenesis*. Eds. C.C. Capen, E. Dybing, J.M.Rice and J.D.Wilbourn. IARC
Sci Publ. No. **147**. Lyon, 225.

ICRP (1975). *Report of the Task Group on Reference Man*. Pergamon Press, New York.

IPCS (1994) Assessing human health risks of chemicals: Derivation of guidance values for health-based exposure limits. Geneva, World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria 170; http://www.inchem.org/documents/ehc/ehc/ehc170.htm)

Kalberlah, F and Schneider, K (1998). Quantification of extrapolation factors. Final report of the research project No 116 06 113 of the Federal Environmental Agency, Germany.

Kroes, R; Renwick, AG; Cheeseman, M et al (2004) Structure based thresholds of toxicological concern (TTC): guidance for substances present at low concentrations in the diet. Fd Chem Toxicol 42: 65-83

MacInnes JI, Nozik ES, Kurtz DT. (1986) Tissue-specific expression of the rat alpha 2u-globulin gene family. Mol Cell Biol 6 (10):3563-3567.

Mancini MA, Majumdar D, Chatterjee B, Roy AK. (1989) Alpha 2u-globulin in modified sebaceous glands with pheromonal functions: localization of the protein and its mRNA in preputial, meibomian, and perianal glands. J Histochem Cytochem 37 (2):149-157.

Makris, SL; Raffaele, K; Allen, S *et al* (2009) A retrospective performance assessment of the developmental neurotoxicity study in support of OECD test guideline 426. *Environmental Health Perspectives* 117(1).

Murty CV, Sarkar FH, Mancini MA, Roy AK. (1987) Sex-independent synthesis of alpha 2u-globulin and its messenger ribonucleic acid in the rat preputial gland: biochemical and immunocytochemical analyses. Endocrinology 121(3):1000-1005.

National Research Council (NRC). (1977). Drinking Water and Health. Washington, DC: National Academies Press.

National Research Council (NRC). (1983). Risk Assessment in the Federal Government: Managing the Process. Washington, DC: National Academy Press.

NRC (2007) National research council of the national academies. Toxicity Testing in the 21st Century: A Vision and a Strategy http://www.nap.edu/catalog/11970.html ISBN: 0-309-10993-0, 216 pages, 6 x 9, (2007)

National Research Council (NRC). (2009). Science and Decisions: Advancing
Risk Assessment. Committee on Improving Risk Analysis Approaches Used by the USEPA Washington, DC: National Academies Press.

Neumann HG. (2009). Risk assessment of chemical carcinogens and thresholds. *Crit Rev Toxicol* 39: 449–461.

OECD (2011). Report of the Workshop on Using Mechanistic Information in Forming Chemical Categories. OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 138. ENV/JM/MONO(2011)8, 2011.

OECD (2012a). Proposal for a template, and guidance on developing and assessing the completeness of adverse outcome pathways.
www.oecd.org/env/ehs/testing/49963554.pdf.

OECD (2012b). The adverse outcome pathway for skin sensitisation initiated by covalent binding to proteins. Part 1: scientific evidence. Series on testing and assessment No. 168. ENV/JM/MONO(2012)10/PART1.

OECD (2012c). The adverse outcome pathway for skin sensitisation initiated by covalent binding to proteins. Part 2: use of the AOP to develop chemical categories and integrated assessment and testing approaches. Series on testing and assessment No. 168. ENV/JM/MONO(2012)10/PART2.

Poulin, P and Krishnan, K (1995). A biologically-based algorithm for predicting human tissue:blood partition coefficients of organic chemicals. *Hum. Exp. Toxicol.* 14, 273-280.

Poulin, P and Krishnan, K (1996). A mechanistic algorithm for predicting blood:air partition coefficients of organic chemicals with the consideration of reversible binding in hemoglobin. *Toxicol. Appl. Pharmacol.* 136: 131-137.

Pratt I, Barlow S, Kleiner J and Larsen JC (2009). The influence of thresholds on the risk assessment of carcinogens in food. *Mutat Res* 678: 113–117.

Rao, GN, Edmondson, J and Elwell, MR (1993). Influence of dietary protein concentration on severity of nephropathy in fischer-344 (F-344/N) rats. *Toxicologic Pathology*, 21: 353-361.

Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC, OJ L 309, 24.11.2009, p. 1-50

Renwick, AG (1993). Data-derived safety factors for the evaluation of food additives and environmental contaminants. *Food Additives and Contaminants*, 10: 275-305.

Renwick AG and Lazarus NR (1998). Human variability and noncancer risk assessment – an analysis of the default uncertainty factor. *Regul Toxicol Pharmacol*, 27: 3–20.

Renwick A.G, Barlow S.M, Hertz- Picciotto I, Boobis A.R, Dybing E, Edler L, Eisenbrand G, Greig JB, Kleiner J, Lambe J, Müller D.J.G, Smith M.R, Tritscher A, Tuijtelaars S, van den Brandt P.A, Walker R, Kroes R, 2003. Risk Characterisation of chemicals in food and diet. Food and Chemical Toxicology, 41, 1211-1271.

Rhomberg, LR and Wolff, SK (1998). Empirical scaling of single oral lethal doses across mammalian species base on a large database. *Risk Anal*. 18: 741-53.

Rhomberg, LR and Lewandowski, TA (2004). Methods for identifying a default cross-species scaling factor. Document prepared for the Risk Assessment Forum, USEPA.

Rhomberg LR, Goodman JE, Haber LH, Dourson M, Andersen ME, Klaunig JE, Meek B, Price PS, McClellan RO and Cohen SM (2011). Linear low-dose extrapolation for noncancer health effects is the exception, not the rule. *Crit Rev Toxicol* 41(1): 1-19.

Ruggenenti, P, Schieppati, A, and Remuzzi, G (2001). Progression, remission, regression of chronic renal disease. *Lancet*, 357: 1601-1608.

Solecki, R; Davies, L; Dellarco, V et al (2005) Guidance on setting of acute reference dose (ARfD) for pesticides. Fd Chem Toxicol 43: 1569 – 1593.

Solecki, R; Moeller, T; Herrman, M et al (2010) A retrospective analysis of acute reference doses for pesticides evaluated in the European Union. Crit Rev Toxicol 40 (1) : 24 – 34.

Turkstra, GH, and van Raaij, MTM (2001). $Alpha_{2u}$-globulin associated nephropathy and renal-cell neoplasms. *RIVM report 601516009, Factsheet FSV-006/00 dated 13-04-2001.* Pages 13-24. *http://www.rivm.nl/bibliotheek/rapporten/601516009.pd*

Schneider K, Oltmanns J and Hassaner M (2004) Allometric principles for interspecies extrapolation in toxicological risk assessment - empirical investigations. *Regulatory Tox Pharm* 39(3): 334-347.

Slob, W and Pieters, MN (1998). A probabilistic approach for deriving acceptable human intake limits and human health risks from toxicological studies: general framework. *Risk Anal.* 18(6): 787-98.

Slob W (1999). Thresholds in toxicology and risk assessment. *Int J Tox*, 18:259-268.

Sonich-Mullin, C, Fielder, R, Wiltse, J, Baetcke, K, Dempsey, J, Fenner-Crisp, P, Grant, D, Hartley, M, Knaap, A, Kroese, D, Mangelsdorf, I, Meek, E, Rice, JM and Younes, M IPCS conceptual

framework for evaluating a mode of action for chemical carcinogenesis. *Reg.Toxicol.Pharmacol.* 34: 146-152, 2001.

Swenberg, JA and Lehman-McKeeman, LD (1999). α_{2u}-globulin-associated nephropathy as a mechanism of renal tubule cell carcinogenesis in male rats. In*: Species Differences in Thyroid, Kidney and Bladder Carcinogenesis*, eds CC Capen, E Dybing, JM Rice and JD Wilbourn. IARC Scientific Publications No. 147, International Agency for Research on Cancer, Lyon, 1999.

Theil, FP, Guentert, TW, Haddad, S, and Poulin, P (2003). Utility of physiologically based pharmacokinetic models to drug development and rational drug discovery candidate selection. *Toxicology Letters* 138: 29-49.

Travis, KZ, Pate, I. and Welsh ZK. (2005). The role of the benchmark dose in a regulatory context. *Regul. Toxicol. Pharmacol*. 43(3): 280-91.

USEPA. (1995). The use of the benchmark dose approach in health risk assessment. Risk Assessment Forum, Office of Research and Development, U.S. Environmental Protection Agency, Washington DC.

USEPA. (1995). The use of the benchmark dose approach in health risk assessment. Risk Assessment Forum, Office of Research and Development, U.S. Environmental Protection Agency, Washington DC.

USEPA (2010). Recommended Toxicity Equivalence Factors (TEFs) for Human Health Risk Assessments of 2,3,7,8Tetrachlorodibenzo-*p*-dioxin and Dioxin-Like Compounds. EPA/100/R-10/005 December 2010. U.S. Environmental Protection Agency, Washington, DC.
http://www.epa.gov/raf/files/tefs-for-dioxin-epa-00-r-10-005-final.pdf

USEPA (2012a). Benchmark Dose Technical Guidance. EPA/100/R-12/001 June 2012. U.S. Environmental Protection Agency, Washington, DC.
http://www.epa.gov/raf/publications/pdfs/benchmark_dose_guidance.pdf

USEPA (2012b). Framework for Human Health Risk Assessment to Inform Decision Making. 601-D12-001 July 12, 2012. U.S. Environmental Protection Agency, Washington, DC.
http://www.epa.gov/raf/files/framework-document-7-13-12.pdf

USEPA (1991). Risk Assessment Forum. Alpha$_{2u}$-globulin: association with chemically induced renal toxicity and neoplasia in the male rat. Report no. EPA/625/3-91/019F.

USEPA (1992). Draft Report: A cross-species scaling factor for carcinogen risk assessment based on equivalence of mg/kg$^{3/4}$/day; Notice. Federal Register 57(109):24152-224173.

USEPA (2000). Supplementary guidance for conducting health risk assessment of chemical mixtures. EPA/630/R-00/002. August. US Environmental Protection Agency, Washington, DC.

USEPA (2002) Hepatocellular hypertrophy. HED guidance document G2002.01, October 21, 2002. Office of Pesticide Programs, USEPA.

USEPA. (2005). Guidelines for Carcinogen Risk Assessment. Washington, DC: Risk Assessment Forum, US Environmental Protection Agency. EPA/630/P- 03/001F.

USEPA (2006). External Review Draft: Harmonisation in interspecies extrapolation – use of $BW^{3/4}$ as default method in derivation of the oral RfD. Risk Assessment Forum Technical Panel. EPA/630/R-06/001.

USEPA (2006): Approaches for the Application of Physiologically-Based Pharmacokinetic (PBPK) Models and Supporting Data in Risk Assessment. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=157668.

van den Berg, M; Birnbaum, L; Bosveld, AT, *et al*. (1998) Toxic equivalency factors (TEFs) for PCBs, PCDDs, PCDFs for humans and wildlife. *Environ Health Perspect* 106(12): 775−792.

van Raaij, MTM (2001). Follicular thyroid tumours in rodents. *RIVM report 601516009, Factsheet FSV-006/00 dated 13-04-2001*. Pages 27-42. *http://www.rivm.nl/bibliotheek/rapporten/601516009.pdf*

VROM. (1989). Memorandum "Dealing with risks" (in Dutch: "Omgaan met risico's"). Appendix no. 5 of the National Plan for the Environment of 1989. Ministry of Housing, Spatial Planning and the Environment. *Lower House of the Dutch Parliament, meeting year 1988/*1989, document 21.137 no. 5. The Hague, The Netherlands: SDU Publishers.

Walton, K, Dorne, JL and Renwick, AG (2001a). Uncertainty factors for chemical risk assessment: interspecies differences in the in vivo pharmacokinetics and metabolism of human CYP1A2 substrates. *Food Chem. Toxicol*. 39: 667-80.

Walton, K, Dorne, JL, and Renwick, AG (2001b). Uncertainty factors for chemical risk assessment: interspecies differences in glucuronidation. *Food Chem. Toxicol*. 39: 1175-90.

Walton, K, Dorne, JL. and Renwick, AG (2004). Species-specific uncertainty factors for compounds eliminated principally by renal excretion in humans. *Food Chem. Toxicol*. 42: 261-74.

White RH, Cote I, Zeise L, Fox M, Dominici F, Burke TA, White PD, Hattis DB, Samet JM (2009). State-of-the-Science Workshop Report: Issues and Approaches in Low Dose–Response Extrapolation for Environmental Health Risk Assessment. Available at: http://www.ehponline.org/ members/2008/ 11502/11502.pdf. Accessed on 3 December 2010.

WHO (1990) *Environmental Health Criteria* 116 . Tributyltin compounds. http://www.inchem.org/documents/ehc/ehc/ehc116.htm.

WHO (2006) Pesticide residues in food 2006. Report of the Joint FAO/WHO meeting on pesticide residues. pp 13 – 17. Food and Agriculture Organisation, Rome. ISBN 978-92-5-105638-7

WHO (2011). WHO guidelines for drinking-water quality, 4th edition. Editor World Health Organisation. [http://www.who.int/water_sanitation_health/dwq/guidelines/en/]

WHO/IPCS (2011). Characterisation and Application of Pharmacokinetic Models in Risk Assessment. Harmonisation Project Document No 9.

WHO (World Health Organization), 1987. International program on Chemical Safety: Principles for the safety assessment of food additives and contaminants in food, Environmental Health Criteria, 70, 174pp,

WHO (World Health Organization), 1994. International Programme on Chemical Safety: assessing human health risks of chemicals: derivation of guidance values for health-based exposure limits. Environmental Health Criteria, 170, 73pp,

WHO (World Health Organization), 1997. Guidelines for predicting dietary intake of pesticide residues, 34 pp.

WHO (World Health Organization), 2001. International Programme on Chemical Safety: guidance document for the use of chemical-specific adjustment factors (CSAFs) for interspecies differences and human variability in dose–concentration response assessment. 76pp,

WHO (World Health Organization), 2005. International Programme on Chemical Safety: Chemical-specific adjustment factors for Interspecies differences and human variability: Guidance document for use of data in dose/concentration–response assessment. IPCS harmonization project document; no. 2. ISBN 92 4 154678 6

Williams, G and Iatropoulos, M(2002). Alteration in liver cell functionand proliferation: differentiation between adaptation and toxicity. *Toxicologic Pathology* 30: 41 – 53.

Zoeller RT, Brown TR, Doan LL, Gore AC, Skakkebaek NE, Soto AM, Woodruff TJ and Vom Saal FS (2012). Mini review: Endocrine-disrupting chemicals and public health protection: a statement of principles from the endocrine society. *Endocrinology* 153(9): 1-14.

ANNEX 1

List of Supporting Excel spreadsheets and description of contents.

1. Master sheet. Contains all the data on ADI, AOEL and ARfD derivations for all active substances, including EPCO, PRAPeR, JMPR and List 4 compounds.

2. PRAPER & EPCO discussions. Contains details of the main discussion topics during the EU peer review of the pesticide active substances.

3. List 4 split. Contains details of ADIs, AOELs and ARfDs for list 4 compounds separately and the overall list minus the list 4 compounds

4. ADI separated info. Contains a number of sub-sheets with compounds separated by different parameters relating to the derivation of the ADIs.

5. AOEL separated info. Contains a number of sub-sheets with compounds separated by different parameters relating to the derivation of the AOELs.

6. ARfD separated info. Contains a number of sub-sheets with compounds separated by different parameters relating to the derivation of the ARfDs.

7. Biocide Pesticide comparison. Contains details for compounds evaluated in the EU with both a biocide AEL and pesticide AOEL values.

8. Liver wt & body wt. Details of the magnitude of changes used in determining LOAELs or NOAELs for compounds with reference doses based on increased liver weight or reduced body weight.

These spreadsheets contain the basic data supporting the analyses presented in the body of the report. The information is presented as searchable files to permit 3rd parties to perform their own analyses with the data.

ANNEX 2

Results of the questionnaire on current approaches to setting health based guidance values.

Due to difficulties presenting information as embedded files or condensed to fit an A4 page and be legible, the details of the questionnaires are presented as supporting excel files.

- A spreadsheet summarising the individual responses (Questionnaire Combined responses summary.xls)

- Individual responses can be made available on request.

ANNEX 3

Detailed Summary of BMD analyses

Active and Reference value	Effect (study)	NOAEL (mg/kg bw)	LOAEL (mg/kg bw)	Effect size at the LOAEL	PROAST 36.5 BMD10 (mg/kg bw)	PROAST 36.5 Lowest BMDL10 – highest BMDU10	PROAST 36.5 Imprecision factor (BMDU/BMDL)	PROAST 36.5 Plausible models	Comment
Carbofuran ADI & ARfD	Brain Cholinesterase in rat pups (acute neurotox study)	<0.03	0.03	↓20%F; ↓13%M	0.017 -0.019	0.006 – 0.037	6	Hills and exponential (continuous variable)	Good data set; No significant sex differences *BMDS BMDLs 0.02 – 0.14*
Carbofuran AOEL	Brain Cholinesterase in rat adults (acute neurotox study)	0.03	0.1	↓20%F; ↓32%M	F: 0.04 -0.05 M: 0.02 - 0.04	0.015 – 0.078	5	Hills and exponential (continuous variable)	Good dataset; males more sensitive. *BMDS BMDLs 0.026 – 0.11*
Cyanamide ADI & AOEL	Decrease in T3 (90-d dog study)	<0.6	0.6	↓16%M; ↓17%F	0.4	0.10 – 1.81	18	Hills and exponential (continuous variable)	Small group size; No significant sex difference *BMDS BMDLs 0.02 – 0.12*
	Decrease in T4 (90-d dog study)	<0.6	0.6	↓32%M; ↓16%F	BMD30 = 0.46 - 0.47 (M);	BMDL30 – BMDU30 = 0.03 – 1.57 (M)	50 (M) 413 (F)	Hills and exponential (continuous variable)	BMD10 very imprecise (BMDL10 = 0) as response at the LOAEL is much greater; Hence,

Active and Reference value	Effect (study)	NOAEL (mg/kg bw)	LOAEL (mg/kg bw)	Effect size at the LOAEL	PROAST 36.5 BMD10 (mg/kg bw)	PROAST 36.5 Lowest BMDL10 – highest BMDU10	PROAST 36.5 Imprecision factor (BMDU/BMDL)	PROAST 36.5 Plausible models	Comment
					BMD30 = 9 - 10 (F)	1.16 – 479 (F)			BMD30 calculated; Males more sensitive; Large CI and imprecision as large scatter in the data. *BMDS BMDLs 0.02 – 0.33*
	Increased aspermatogenesis (90-d dog study)	<0.6	0.6	↑50%	NA	0 – 2.54	Infinite	LVM E5 LVM H2 Weibull Log-prob Two-stage Gamma Log-logistic (quantal variable)	BMDL10 = 0, hence PoD cannot be derived; Small group size. *BMDS BMDLs 0.07 – 0.48*
Cyanamide ARfD	Maternal hypoactivity (rat develop study)	5	15	↑32%	7.12 – 9.25	0.9 – 14.6	16	Two-stage Log-logistic Weibull Log-prob Gamma LVM E4 LVM H5 (quantal variable)	Imprecision factor relatively high, but not too bad given the data *BMDS BMDLs 2.9 – 15.6*

Active and Reference value	Effect (study)	NOAEL (mg/kg bw)	LOAEL (mg/kg bw)	Effect size at the LOAEL	PROAST 36.5 BMD10 (mg/kg bw)	PROAST 36.5 Lowest BMDL10 – highest BMDU10	PROAST 36.5 Imprecision factor (BMDU/BMDL)	PROAST 36.5 Plausible models	Comment
Etridiazole ADI	Increased kidney cell karyomegaly (2-yr rat study)	<5	5	↑62%M ↑78%F	0.03 – 2.4	0.0005 – 3.63	7260	Log-logistic Weibull Log-prob Gamma LVM E5 LVM H2 (quantal variable)	BMD10 CI are very wide (4 orders of magnitude) as the lowest dose is too high; High imprecision; data not suitable to derive PoD. *BMDS BMDLs* *0.47 – 2.0*
Etridiazole AOEL	Increased ALP (1-yr dog study)	3.1	8.07	↑97%M ↑88%F	BMD50 = 2.85 – 2.97	BMDL50-BMDU50 = 1.09 – 4.5	4	Hills and exponential (continuous variable)	BMD10 very imprecise as a very large maximum response, hence BMD50 calculated; No sex differences. *BMDS BMDLs* *0.2 – 1.1*
	Increased liver wt (1-yr dog study)	3.1	8.07	↑12%M	BMD05 = 3.29 – 3.38	BMDL05 – BMDU05 = 0.68 – 7.89	10	Hills and exponential (continuous variable)	BMD05 calculated; No significant sex differences; Reasonable precision given the small group sizes *BMDS BMDLs* *0.2 – 0.46*
Fenoxycarb ADI	Lung tumours (18-mth mouse	<5.3	5.3	↑10%M	NA	3.29×10^{-6} – 64.3	Very large	One-stage Logistic	Very large imprecision owing to high

Active and Reference value	Effect (study)	NOAEL (mg/kg bw)	LOAEL (mg/kg bw)	Effect size at the LOAEL	PROAST 36.5 BMD10 (mg/kg bw)	PROAST 36.5 Lowest BMDL10 – highest BMDU10	PROAST 36.5 Imprecision factor (BMDU/BMDL)	PROAST 36.5 Plausible models	Comment
	study)							Log-logistic Weibull Probit Log-probit Gamma LVM E2 LVM H2 (quantal variable)	background rate and small increase in response; Data not suitable to derive PoD. *BMDS BMDLs* *11 – 17*
Fenoxycarb AOEL	Increased liver wt (90-d rat study)	9.7	45.1	↑16%M ↑17%F	BMD05 = 21.5 – 22.4 (M) BMD05 = 11.6 – 12.3 (F)	BMDL05 – BMDU05 = 8.6 – 42.4 (M) BMDL05 – BMDU05 = 5.4 – 21.0 (F)	5 (M) 4 (F)	Hills and exponential (continuous variable)	BMD05 calculated; F more sensitive; Good precision. *BMDS BMDLs* *20 - 46*
	Increased thyroid wt (90-d rat study)	9.7	45.1	↑27%F	BMD05 = 41.8 – 42.5	BMDL05 – BMDU05 = 0.22 – 163.3	740	Hills and exponential (continuous variable)	BMD05 calculated; High imprecision; No significant sex differences. *BMDS BMDLs* *8.2 - 98*
	Liver hypertrophy (90-d rat study)	9.7	45.1	↑80%F	4.47 - 39	2.65 – 45.1 (F)	20 (F)	One-stage Logistic Log-logistic Weibull Probit Log-probit	Effect seen only in F; Medium imprecision *BMDS BMDLs* *2.7 – 14.5*

Active and Reference value	Effect (study)	NOAEL (mg/kg bw)	LOAEL (mg/kg bw)	Effect size at the LOAEL	PROAST 36.5 BMD10 (mg/kg bw)	PROAST 36.5 Lowest BMDL10 – highest BMDU10	PROAST 36.5 Imprecision factor (BMDU/BMDL)	PROAST 36.5 Plausible models	Comment
								Gamma LVM E2 LVM H3 (quantal variable)	
	Thyroid hypertrophy (90-d rat study)	9.7	45.1	↑70%F	5.28 – 39.7 (F)	3.15 – 44.7 (F)	15 (F)	One-stage Two-stage Logistic Log-logistic Weibull Probit Log-probit Gamma LVM E2 LVM H2 (quantal variable)	Effect seen only in F; Medium imprecision *BMDS BMDLs* *3.1 – 15.9*
Fluxapyroxad ADI	Liver hypertrophy (2-yr rat study)	2.1	11	↑58%M ↑56%F	4.42 – 4.72	3.33 – 7.01	2	LVM E4 LVM H5 (quantal variable)	Good precision; No significant sex difference. *BMDS BMDLs* *9 - 27*
	Liver tumours (2-yr rat study)	2.1	11	↑8%M 0% F	28.1 – 59.8	12.3 – 71.4	6	Two-stage Log-logistic Weibull Log-probit	Good precision; Males more sensitive *BMDS BMDLs*

Active and Reference value	Effect (study)	NOAEL (mg/kg bw)	LOAEL (mg/kg bw)	Effect size at the LOAEL	PROAST 36.5 BMD10 (mg/kg bw)	PROAST 36.5 Lowest BMDL10 – highest BMDU10	PROAST 36.5 Imprecision factor (BMDU/BMDL)	PROAST 36.5 Plausible models	Comment
								Gamma LVM E2 LVM H2 (quantal variable)	11 - 21
	Liver spongiosis (2-yr rat study)	2.1	11	↑20%M 0%F	28.3 – 40.7	21.1 – 59.2	3	Two-stage Log-logistic Weibull Log-probit Gamma LVM H2 (quantal variable)	Good precision; Males more sensitive *BMDS BMDLs* *0.08 – 7.9*
	Tooth whitening (2-yr rat study)	2.1	11	↑64%M ↑72%F	18.7 – 27.8	14.1 – 23.9	1.7	Log-logistic Log-probit LVM E4 LVM H5 (quantal variable)	Very good precision; No significant sex differences *BMDS BMDLs* *19 - 49*
Fluxapyroxad AOEL	Increased T3 (90-d rat study)	6.0	31	↑22%M ↑10%F	BMD30 = 63.1 – 64.8	BMDL30 – BMDU30 = 4.5 - 1107	246	Hills and exponential (continuous variable)	BMD10 not appropriate (BMDL10 = 0) as the effect at the LOAEL is much higher, hence BMD30 calculated; No significant sex difference;

Active and Reference value	Effect (study)	NOAEL (mg/kg bw)	LOAEL (mg/kg bw)	Effect size at the LOAEL	PROAST 36.5 BMD10 (mg/kg bw)	PROAST 36.5 Lowest BMDL10 – highest BMDU10	PROAST 36.5 Imprecision factor (BMDU/BMDL)	PROAST 36.5 Plausible models	Comment
									Large imprecision; Not suitable to derive PoD. *BMDS BMDLs 3.5 - 201*
	Increased T4 (90-d rat study)	6.0	31	↑26%F No effect in M at LOAEL	10^{-6}	CI infinitely large	Infinitely large	None	Treated groups show response that are not significantly increasing; The controls differ from treated for reasons other than the dose; Not suitable to derive PoD. *BMDS BMDLs 0.00 – 1.3*
	Decreased clotting time (90-d rat study)	6.0	31	↓8%F No effect in M at LOAEL	34	24.2 – 96.8	4	Hills and exponential (continuous variable)	Good precision; No significant sex differences. *BMDS BMDLs 37 - 49*
Fluxapyroxad ARfD	Decreased maternal bwt (Rat dev study)	25	200	↓25%	BMD15 = 24	BMDL15 – BMDLU15 = 0.76 - 224	300	Hills and exponential (continuous variable) with litter effects	BMD05 beyond range of observations, hence BMD15 calculated; Large imprecision owing to large within-group variance *BMDS BMDLs 20 - 445*
	Resorptions (Rabbit dev study)	25	60	↑15%	66.5 – 76.8	25.3 - Infinite	Infinite	Two-stage Log-logistic	Large imprecision; Note there are 2 nests

Active and Reference value	Effect (study)	NOAEL (mg/kg bw)	LOAEL (mg/kg bw)	Effect size at the LOAEL	PROAST 36.5 BMD10 (mg/kg bw)	PROAST 36.5 Lowest BMDL10 – highest BMDU10	PROAST 36.5 Imprecision factor (BMDU/BMDL)	PROAST 36.5 Plausible models	Comment
								Weibull Log-probit LVM H2 (quantal variable) With litter effects	with 100% response, at the two highest doses- if these were excluded, response would have not been significant. *BMDS BMDLs* *Nested model did not run*
Propanil ADI & AOEL	Decreased RBC (1-yr dog study)	<5	5	↓9%M ↓3%F	BMD05 = 50 – 51.9 (F) BMD05 = 27.7 – 29 (M)	BMDL05 – BMDU05 = 20.7 – 77.5 (F) BMDL05 – BMDU05 = 6.8 – 45.6 (M)	4 (F) 7 (M)	Hills and exponential (continuous variable)	BMD05 estimated; Males more sensitive; Good precision *BMDS BMDLs* *0.2 – 2.5*
	Increased metHb (1-yr dog study)	<5	5	↑0.8%M ↑0.9%F	BMD2X = 21.04 (F)	BMDL2X - BMDU2X = 7.7 – 46 (F)	6	Power model as response in controls is 0 (background response added to dose-response curve rather than being multiplied)	BMD defined as twice the response in controls; Females more sensitive; Good precision *BMDS BMDLs* *0.2 – 1.1*
	Decreased Hb	<5	5	↓9%M	BMD05 =	BMDL05 –	4 (M)	Hills and	BMD05 estimated;

Active and Reference value	Effect (study)	NOAEL (mg/kg bw)	LOAEL (mg/kg bw)	Effect size at the LOAEL	PROAST 36.5 BMD10 (mg/kg bw)	PROAST 36.5 Lowest BMDL10 – highest BMDU10	PROAST 36.5 Imprecision factor (BMDU/BMDL)	PROAST 36.5 Plausible models	Comment
	(1-yr dog study)			↓1.2%F	36.4 – 36.9 (M) BMD05 = 74.6 – 75.1 (F)	BMDU05 = 14.7 – 58.8 (M) BMDL05 – BMDU05 = 46.4 – 226 (F)	5 (F)	exponential (continuous variable)	Males more sensitive; Good precision. *BMDS BMDLs 0.2 – 2.6*
	Increased liver hemosiderin (1-yr dog study)	<5	5	↑25%M 0%F (100% at 40)	2.68 – 12.2	0.4 – 24.9	20	One-stage Two-stage Log-logistic Weibull Log-probit Gamma Logistic Probit LVM E2 LVM H2 (quantal variable)	Females more sensitive; Moderate precision.
	Increased kidney hemosiderin (1-yr dog study)	<5	5	↑100%M ↑25%F	$1X10^{-6}$ – 1.89	0 - 5	Infinite	One-stage Two-stage Log-logistic Weibull Log-probit Gamma Logistic Probit LVM E2	Males more sensitive; Data not suitable to derive PoD as BMDL = 0 as 100% at the lowest dose in M.

Active and Reference value	Effect (study)	NOAEL (mg/kg bw)	LOAEL (mg/kg bw)	Effect size at the LOAEL	PROAST 36.5 BMD10 (mg/kg bw)	PROAST 36.5 Lowest BMDL10 – highest BMDU10	PROAST 36.5 Imprecision factor (BMDU/BMDL)	PROAST 36.5 Plausible models	Comment
								LVM H2 (quantal variable)	
Propanil ARfD	Decreased Hb (30-d dog study)	7	17	↓7%M ↓19%F	BMD05 = 11.8 – 15.7	BMDL05 – BMDU05 = 6.3 – 31.73	5	Hills and exponential (continuous variable)	BMD05 calculated; Sex difference is uncertain; Reasonable precision. *BMDS BMDLs 0.45 – 0.55*
	Decreased MCHC (30-d dog study)	7	17	↓6%M ↓5%F	BMD05 = 15.9	BMDL05 – BMDU05 = 8.5 – 46.6	5	Hills and exponential (continuous variable)	BMD05 calculated; No significant sex differences; Good precision. *BMDS BMDLs 0.02 – 0.42*
	Decreased RBC (30-d dog study)	7	17	↓7%M 19%F	BMD05 = 12.2 – 12.4 (F) BMD05 = 16.6 (M)	BMDL05 – BMDU05 = 8.2 – 13.2 (F) BMDL05 – BMDU05 = 0 – NA (M)	1.6 (F) Infinite (M)	Hills and exponential (continuous variable)	BMD05 calculated; Females more sensitive; For males BMDL05 = 0, so PoD for males cannot be derived. Good precision for F. *BMDS BMDL1.0*
Sulcotrione ADI	Kidney enlargement (2-yr rat study)	<0.04	0.04	↑12%M	NA	0 – 0.39 (M)	Infinite	Two-stage Log-logistic Weibull Log-probit	Very large imprecision. BMDL10 = 0, so PoD cannot be derived. *BMDS BMDLs*

Active and Reference value	Effect (study)	NOAEL (mg/kg bw)	LOAEL (mg/kg bw)	Effect size at the LOAEL	PROAST 36.5 BMD10 (mg/kg bw)	PROAST 36.5 Lowest BMDL10 – highest BMDU10	PROAST 36.5 Imprecision factor (BMDU/BMDL)	PROAST 36.5 Plausible models	Comment
								Gamma LVM E5 (quantal variable)	*0.017 – 0.83*
	Kidney cystic change (2-yr rat study)	<0.04	0.04	↑12%M	NA	0 – 0.74 (M)	Infinite	One-stage Two-stage Log-logistic Weibull Log-probit Gamma Logistic LVM E5 LVM H2 (quantal variable)	Very large imprecision. BMDL10 = 0, so PoD cannot be derived. *BMDS BMDLs* *0.08 – 0.32*
	Kidney pelvic dilatation (2-yr rat study)	<0.04	0.04	↑14%M	NA	0 – 0.415 (M)	Infinite	Log-logistic Log-probit Gamma LVM E5 LVM H5 (quantal variable)	Very large imprecision. BMDL10 = 0, so PoD cannot be derived. *BMDS BMDLs* *0.12 – 0.93*
Sulcotrione AOEL	Nephropathy in P0 M (2-gen rat study)	0.06	0.6	↑36%M	NA	0.0115 – 0.306 (M)	26	LVM E4 LVM H4 (quantal variable)	Moderate imprecision *BMDS BMDLs* *0.05 – 3.1*

Active and Reference value	Effect (study)	NOAEL (mg/kg bw)	LOAEL (mg/kg bw)	Effect size at the LOAEL	PROAST 36.5 BMD10 (mg/kg bw)	PROAST 36.5 Lowest BMDL10 – highest BMDU10	PROAST 36.5 Imprecision factor (BMDU/BMDL)	PROAST 36.5 Plausible models	Comment
Triazoxide ADI	Increased spleen wt (2-yr rat study)	<0.05	0.05	↑16%F	BMD05 = 3.29 – 3.36	BMDL05 – BMDU05 = 2.6 – 4.2	1.6	Hills and exponential (continuous variable)	BMD05 calculated; Good precision; No sex difference. *BMDS BMDLs 0.00 – 0.27*
	Dark spleen (2-yr rat study)	<0.05	0.05	↑22%M ↑12%F	0.00032 – 0.0016	0 – 0.0287	Infinite	Log-logist Weibull Log-probit Gamma LVM E5 LVM H5 (quantal variable)	Large imprecision. BMDL10 = 0, so PoD cannot be derived. *BMDS BMDLs* 0.04 – 0.65
Triazoxide AOEL	Increased liver wt (90-d rat study)	0.2	2.6	↑12%M&F	BMD05 = 1.45 – 1.82 (M) BMD05 = 0.29 (F)	BMDL05 – BMDU05 = 0.28 – 2.16 (M) BMDL05 – BMDU05 = 0.20 – 0.52 (F)	7.7 (M) 2.6 (F)	Hills and exponential (continuous variable)	BMD05 calculated; Females more sensitive; Good precision. *BMDS BMDLs 0.06 – 2.3*
	Increased spleen wt (90-d rat study)	0.2	2.6	↑9%F	BMD05 = 3.29 – 3.36	BMDL05 – BMDU05 = 2.63 – 4.16	1.6	Hills and exponential (continuous variable)	BMD05 calculated; No sex difference; Very good precision. *BMDS BMDLs 0.35 – 2.6*
	Increased	0.2	2.6	↑80%M	BMD50 =	BMDL50 –	11 (M)	Hills and	As maximum response

Active and Reference value	Effect (study)	NOAEL (mg/kg bw)	LOAEL (mg/kg bw)	Effect size at the LOAEL	PROAST 36.5 BMD10 (mg/kg bw)	PROAST 36.5 Lowest BMDL10 – highest BMDU10	PROAST 36.5 Imprecision factor (BMDU/BMDL)	PROAST 36.5 Plausible models	Comment
	reticulocytes (90-d rat study)			↑54%F	1.3 (M) BMD50 = 3.5 (F)	BMDU50 = 0.27 – 3.0 (M) BMDL50 – BMDU50 = 0.85 – 8.1 (F)	9.5 (F)	exponential (continuous variable)	very large, BMD50 calculated; Males are more sensitive; Moderate imprecision. *BMDS BMDLs* *0.26 – 1.5*
	Spleen congestion (90-d rat study)	0.2	2.6	↑11%M	2.56 – 20.1 (M)	1.67 – 7.94 (M)	5	Log-logist Weibull Log-probit Gamma Logistic Probit LVM E2 LVM H3 (quantal variable)	Good precision; Effect in males only; *BMDS BMDLs* *1.3 – 6.6*

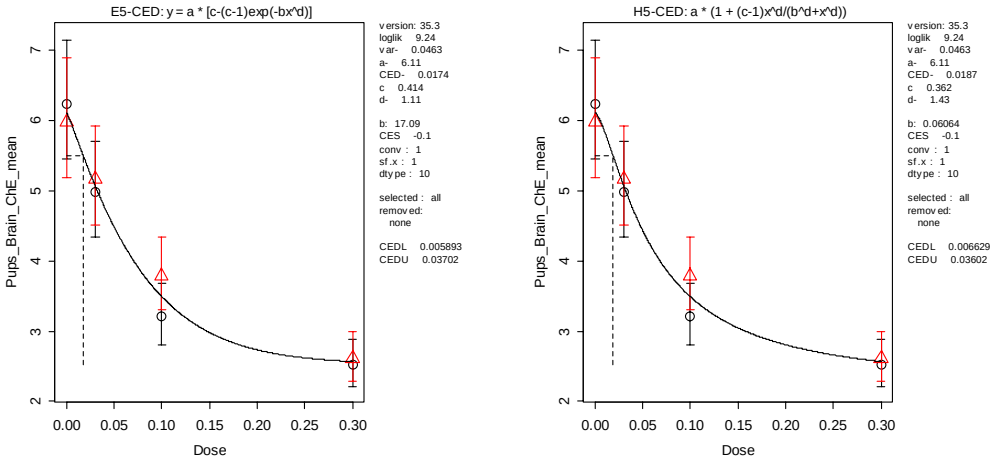
ANNEX 4:

Results of the BMD analyses on 8 pesticides

These analyses were undertaken by Professor W. Slob of RIVM on the PROAST software (http://www.rivm.nl/en/Library/Scientific/Models/PROAST). The comments given are those of Professor Slob. The BMD confidence intervals are the lowest BMDL and the highest BMDU at the set BMR for all the plausible curve fits.

Carbofuran

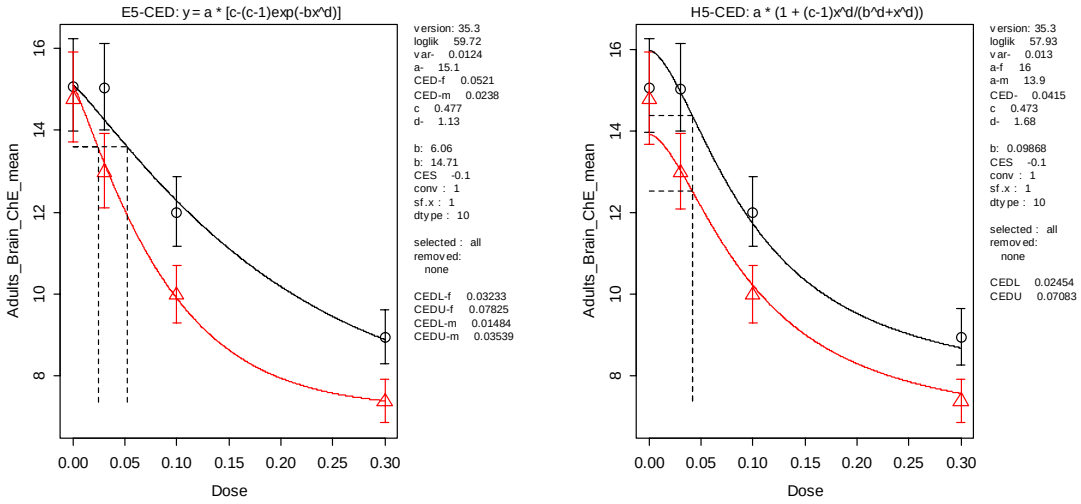
Brain cholinesterase in rat pups



BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% (decrease)	0.006 - 0.037	6

Comments
There no significant sex differences. The imprecision factor of around 6 indicates good data.

Brain cholinesterase in adult rats (acute neurotoxicity study)



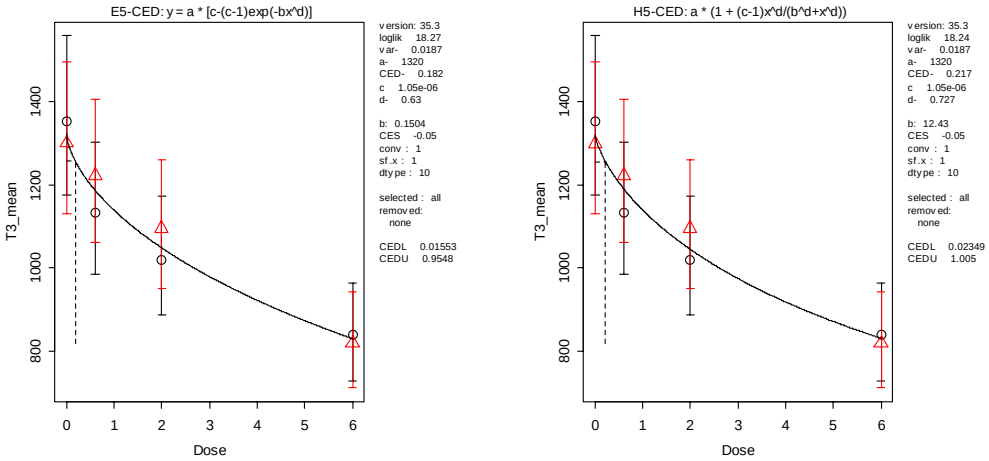
BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% (decrease)	0.015 - 0.078	5

Comments

There is uncertainty in how the sexes differ: in background (right plot) or in sensitivity (left panel). If the latter, females are more sensitive. The imprecision factor is around 5, indicating good data (e.g. good dose location).

Cyanamide

Decrease in T3 (90-day dog study)

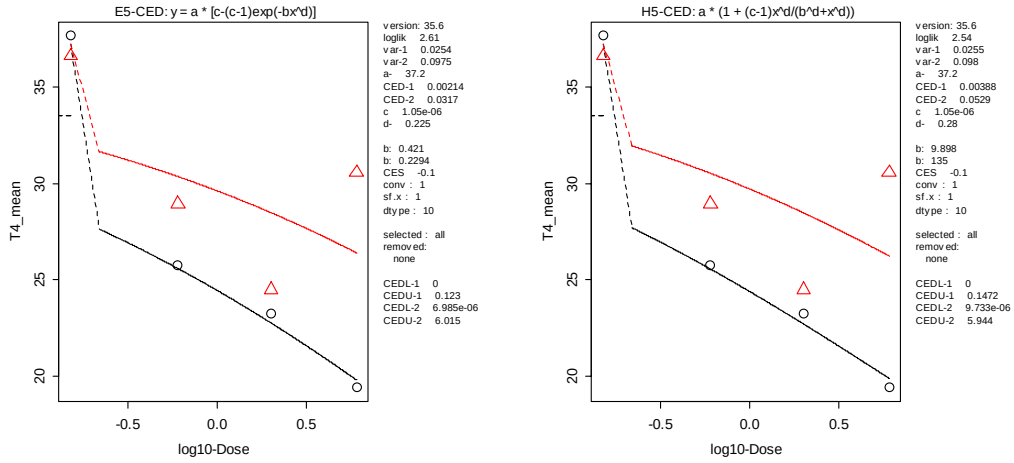


BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
5% (decrease)	0.015 – 1.00	65
10% (decrease)	0.10 – 1.81	18

Comments

At the BMR of 5%, the imprecision factor was 65 (very high), even though the data were good. However, the first dose was a bit too high, and the BMD was left too much freedom. Re-analysis with a BMR of 10% resulted in a better imprecision factor (18).

Decrease in T4 (90-day dog study)

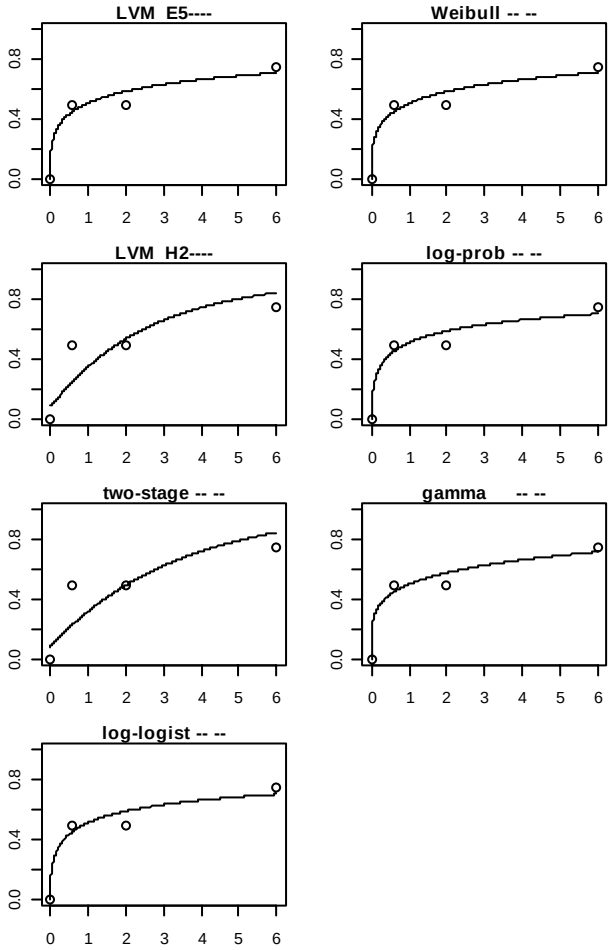


BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% (decrease)	BMDL is 0	Infinite
30% (decrease)	0.03 – 1.57 (males)	50 (males)
	1.16 – 479 (females)	413 (females)

Comments

At a BMR of 10%, the BMDL was zero, owing to the response in the lowest dose being much greater than this. The imprecision factor was around 50 for sex 1 (males – red). For sex 2 it was even larger, which was consistent with the large scatter in the data.

Increased aspermatogenesis (90-day dog study)

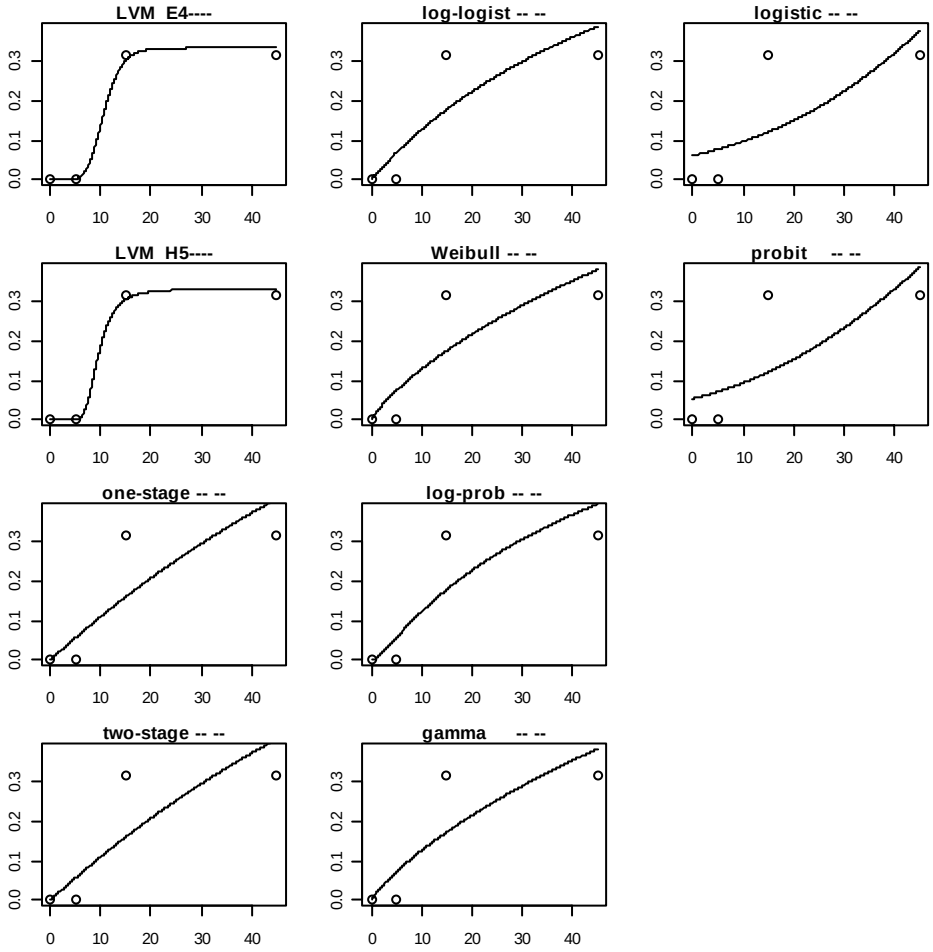


BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% extra risk	0 – 2.54	Infinite

Comments
As the BMDL was 0, these quantal data do not allow the derivation of a PoD. There were only 16 animals in total in the study.

Maternal hypoactivity (rat developmental study)

model	Npar	loglik	accept	BMD	BMDL	BMDU
null	1	-43.97	--	NA	NA	NA
full	4	-31.34	--	NA	NA	NA
two-stage	3	-35.16	yes	9.01	6.11	14
log-logist	3	-34.84	yes	7.46	1.4	14.2
Weibull	3	-35.02	yes	7.12	1.04	14.4
log-prob	3	-34.58	yes	8.08	2.03	14.2
gamma	3	-35.06	yes	7.34	0.912	14.6
LVM: E4-	3	-31.44	yes	9.25	5.65	12.8
LVM: H5-	4	-31.38	yes	8.46	5.56	12.9



BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% extra risk	0.91 – 14.6	16

Comments

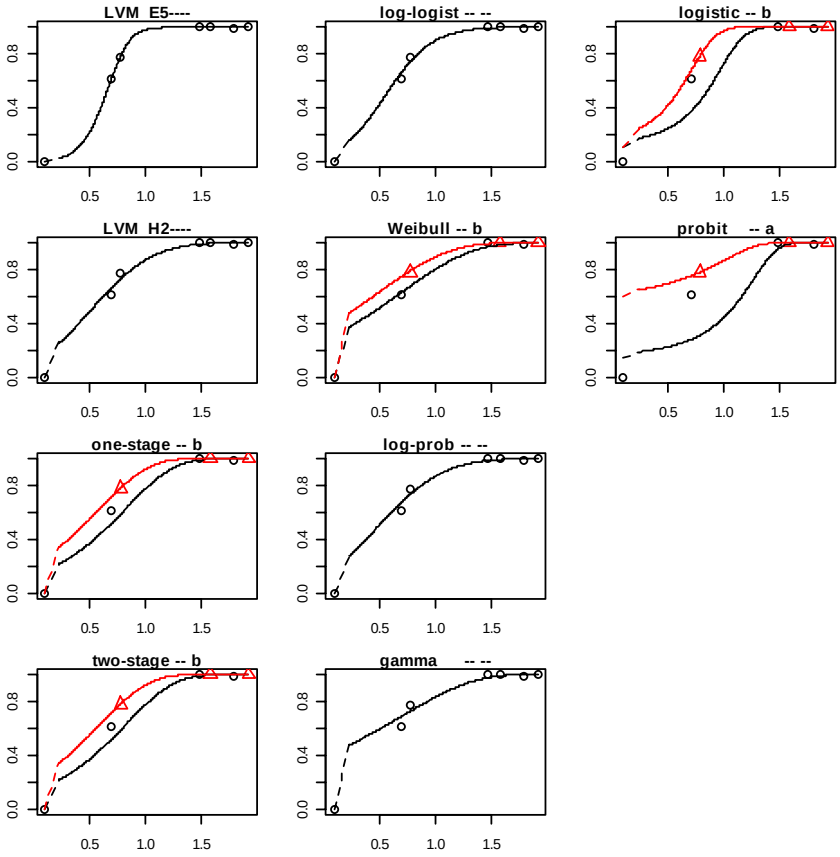
Only the latent variable models in PROAST were accepted by the Goodness of Fit test with P=0.05. However, it is unlikely that the response levels off at 30%. Therefore the P-value is lowered to 0.005, then all models are accepted. Note that this is a conservative approach since we take the lowest BMDL of all accepted models. The imprecision factor is around 16, relatively high, but not too bad given the data.

Etridazole

Kidney cell karyomegaly (2-year rat study)

model	covar	npar	loglik	accept	BMD	BMDL	BMDU	level
null	NA	1	-252.96	--	NA	NA	NA	--
full	NA	7	-64.45	--	NA	NA	NA	--
one-stage	b	3	-70.43	no	0.417	NA	NA	2
two-stage	b	4	-70.43	no	0.417	NA	NA	2
log-logist	--	3	-67.07	yes	1.3	0.609	2.16	1
Weibull	b	4	-67.3	yes	0.121	0.0267	0.319	2

log-prob	--	3	-67.68	yes	0.806	0.336	1.42	1
gamma	--	3	-68.92	yes	0.0305	0.00052	0.245	1
logistic	b	3	-95.17	no	1.23	NA	NA	2
probit	a	3	-104.08	no	1.19	NA	NA	2
LVM: E5-	--	4	-65.85	yes	2.42	0.444	3.63	-
LVM: H2-	--	2	-67.72	yes	0.882	0.592	1.08	-

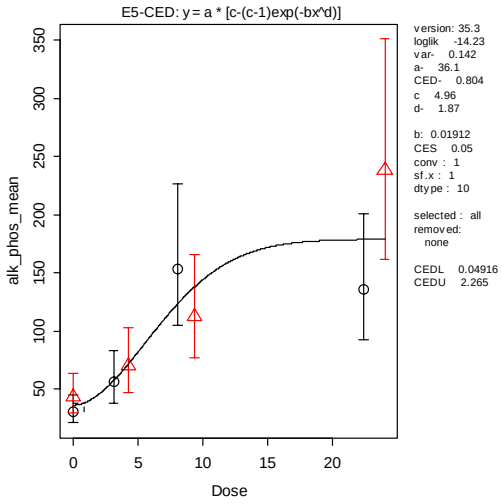


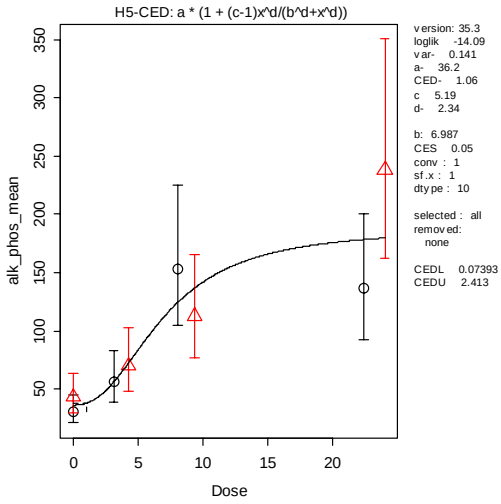
BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% extra risk	0.0005 – 3.63	7260

Comments

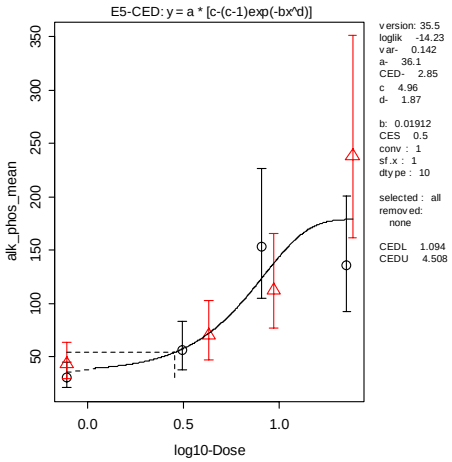
BMD confidence interval is very wide (4 orders of magnitude), owing to the lowest dose being too high. Using the BMDL might nonetheless work, as long as exposure is much lower than the associated exposure limit. Otherwise, the data are not useful for deriving a PoD.

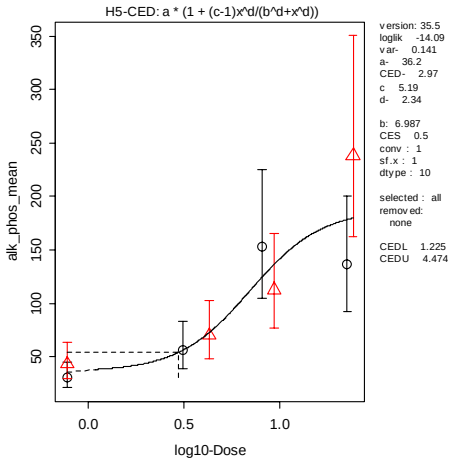
Increased alkaline phosphatase (one-year dog study)





BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
5% (increase)	0.0175 – 4.308	246



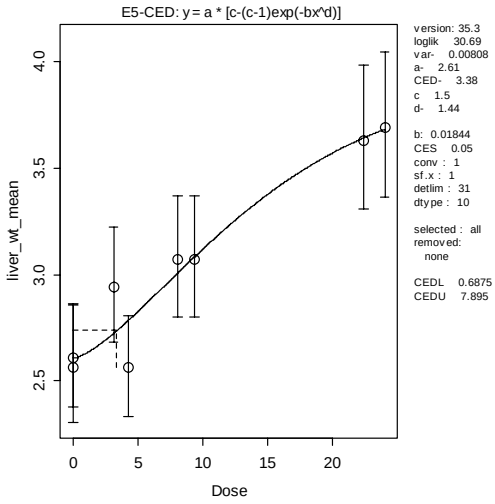


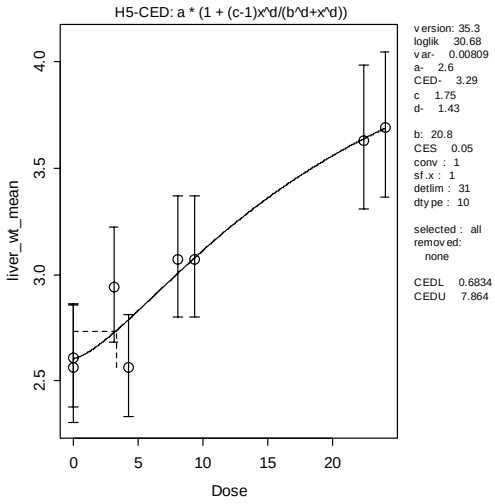
BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
50% (increase)	1.09 – 4.51	4

Comments

With a BMR of 5% the imprecision factor is very large (BMD lower than the lowest dose). Given the large maximum response, a higher than 5% BMR may be used, e.g. 50%. Note the log-dose scale in the last plots. There were no significant sex differences.

Increased liver weight



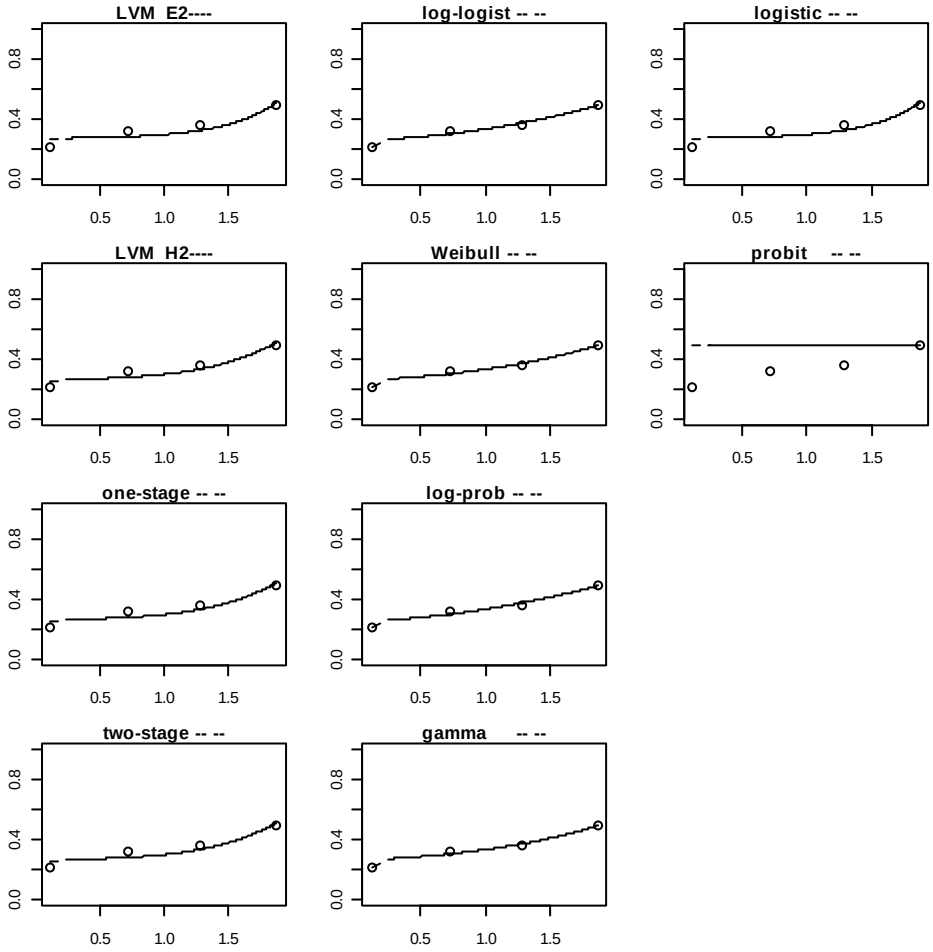


BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
5% (increase)	0.68 – 7.89	10

Comments
There were no significant sex differences.

Fenoxycarb

Lung tumours (18-month mouse study)

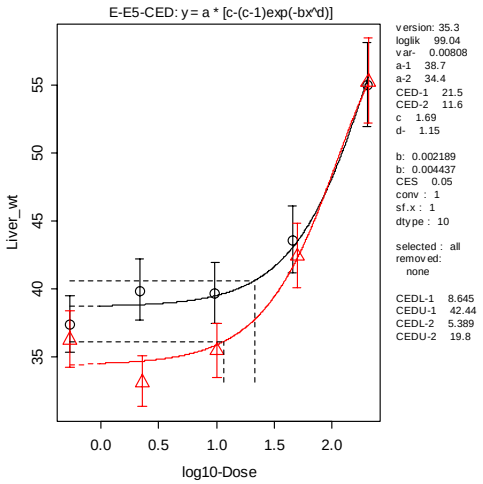


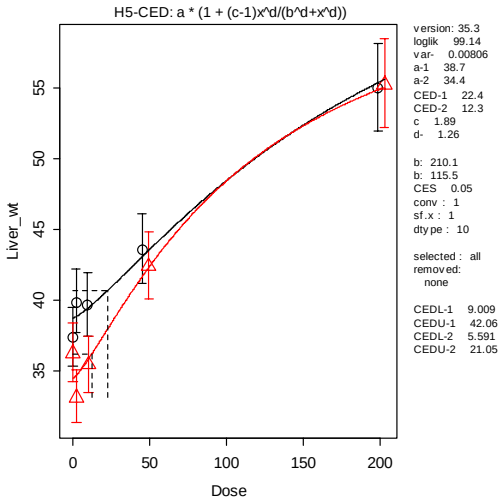
BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% extra risk	3.29×10^{-6} - 64.3	Very large

Comments

There was a very large imprecision factor, owing to weak data: high background and small increase in response. The data are not suitable for the derivation of a PoD.

Increased liver weight (90-day rat study)



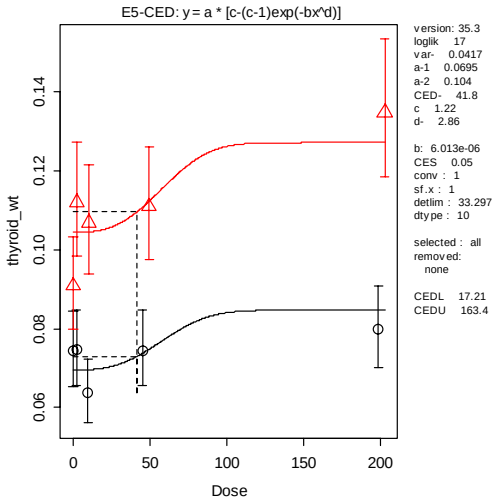


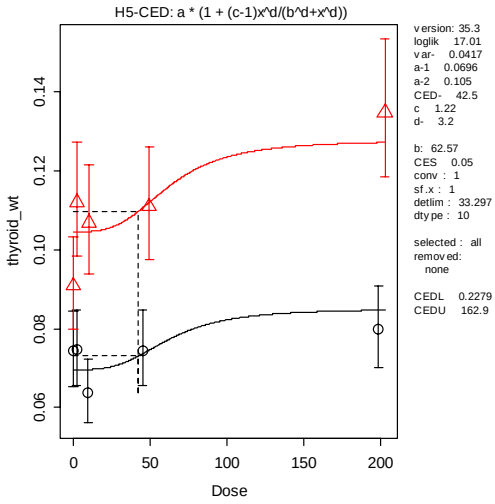
BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
5% (increase)	8.645 – 42.445 (males) 5.389 – 21.049 (females)	5 (males) 4 (females)

Comments

The plot on the left was plotted on a log-dose scale so that the points could be more clearly seen. The females were more sensitive.

Increased thyroid weight (90-day rat study)

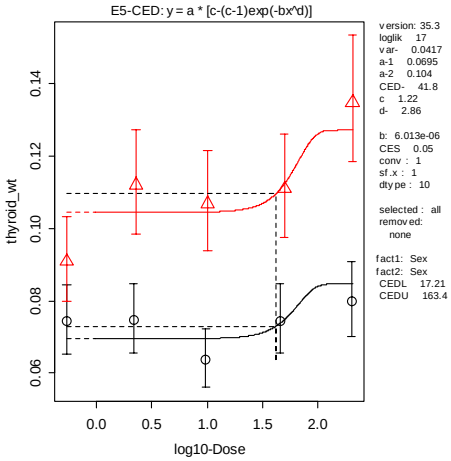




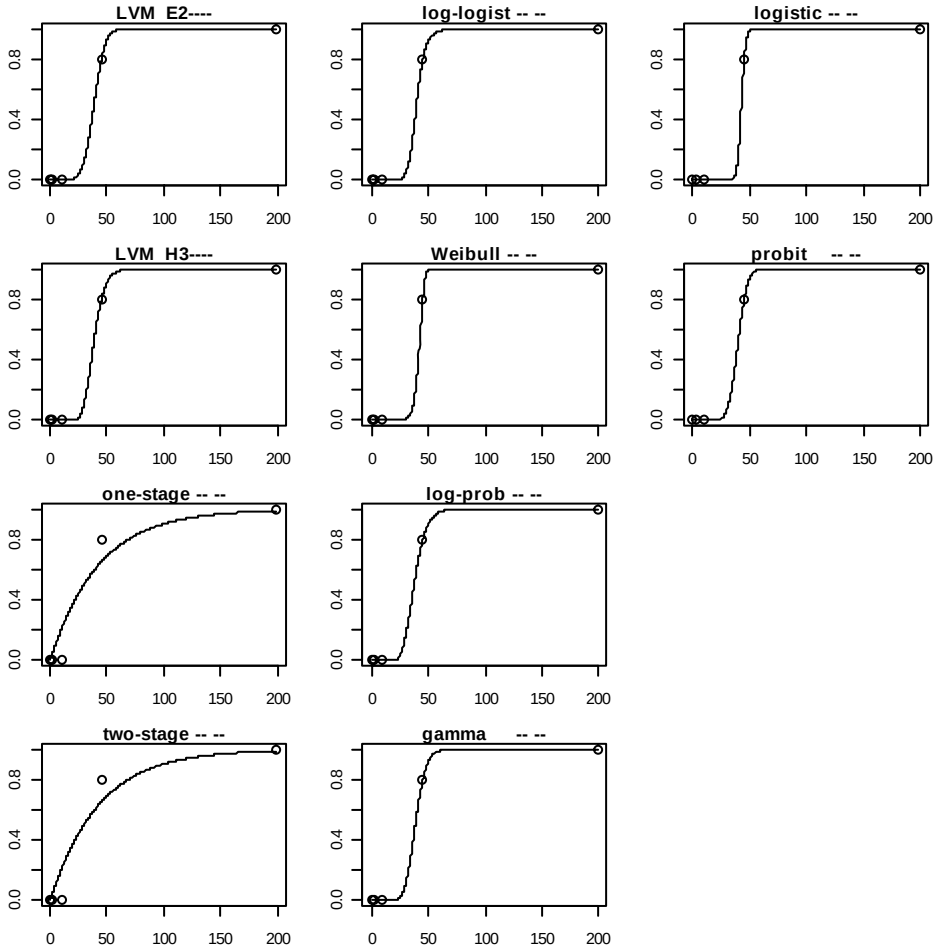
BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
5% (increase)	0.228 – 163.36	740

Comments

The significance of the trend in this data is fully determined by the male response at 200 (see plot on log-dose scale below). It might be that this response is an upwards outlier, just like the controls are a downward outlier. The trend in the females alone would not be significant.



Liver hypertrophy (90-day rat study)

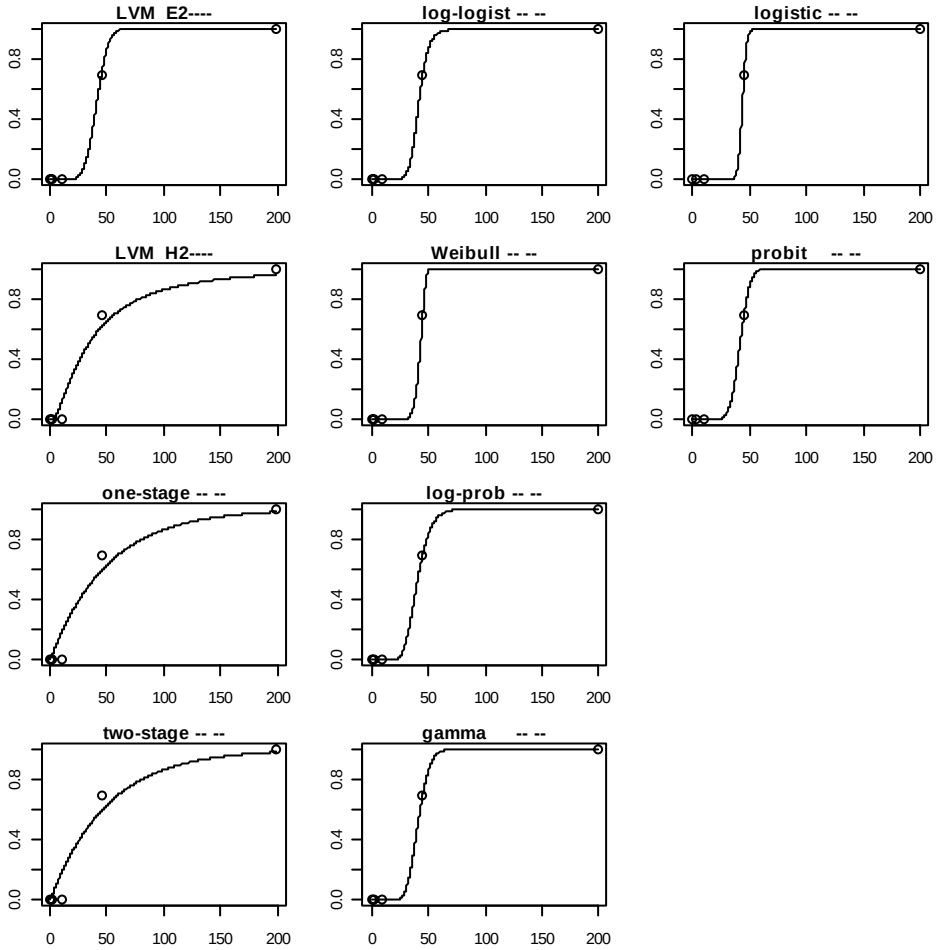


model	npar	loglik	accept	BMD	BMDL	BMDU
null	1	-32.67	--	NA	NA	NA
full	5	-5	--	NA	NA	NA
one-stage	2	-8.41	yes	4.47	2.65	7.73
two-stage	3	-8.41	no	4.47	NA	NA
log-logist	3	-5	yes	32.2	9.37	45.1
Weibull	3	-5	yes	36.5	8.76	38
log-prob	3	-5	yes	28	8.84	NA
gamma	3	-5	yes	29.6	8.72	37.2
logistic	2	-5	yes	39	14.5	NA
probit	2	-5	yes	31.1	15.3	31.2
LVM: E2-	2	-5	yes	28.7	13	32.7
LVM: H3-	3	-5	yes	29.4	8.83	33.9

BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% extra risk	2.65 – 45.1 (females)	20 (females)

Comments
The effects were seen only in females.

Thyroid hypertrophy (90-day rat study)



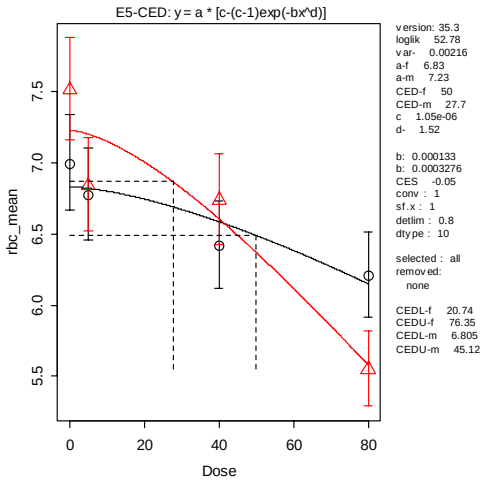
model	npar	loglik	accept	BMD	BMDL	BMDU
null	1	-32.05	--	NA	NA	NA
full	5	-6.11	--	NA	NA	NA
one-stage	2	-8.91	yes	5.28	3.15	9.06
two-stage	3	-8.91	yes	5.28	3.15	9.06
log-logist	3	-6.11	yes	33.4	9.49	44.7
Weibull	3	-6.11	yes	37.2	8.79	NA
log-prob	3	-6.11	yes	29.4	9.05	35.7
gamma	3	-6.11	yes	31.2	8.74	39.7
logistic	2	-6.11	yes	39.7	15.9	NA
probit	2	-6.11	yes	32.6	15.8	NA
LVM: E2-	2	-6.11	yes	30.4	14.1	34.5
LVM: H2-	2	-7.86	yes	9.12	5.23	15

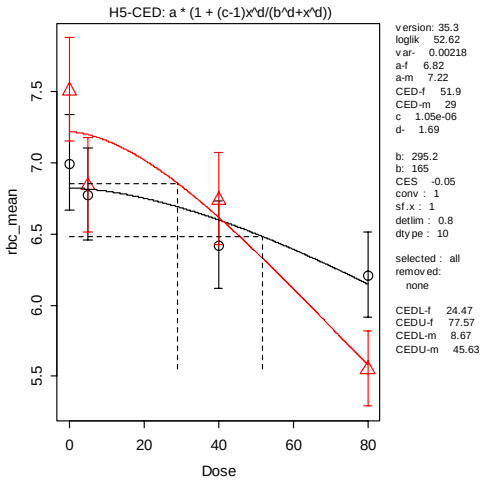
BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% extra risk	3.15 – 44.7 (females)	15 (females)

Comments
The effect was seen only in females.

Propanil

Decreased red blood cells (one-year dog study)



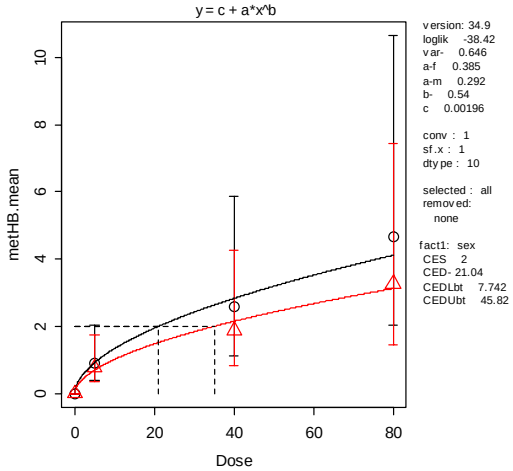


BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
5% (decrease)	6.805 – 45.626 (males)	7 (males)
	20.741 – 77.566 (females)	4 (females)

Comments

There was a significant sex difference, with males appearing to be more sensitive than females (males = red curves).

Increased methaemoglobin (one-year dog study)



BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
2x	7.7 – 46 (females)	6 (females)

Comments

Zero means in controls are replaced by 0.01. The zero SDs are replaced by 0.05.

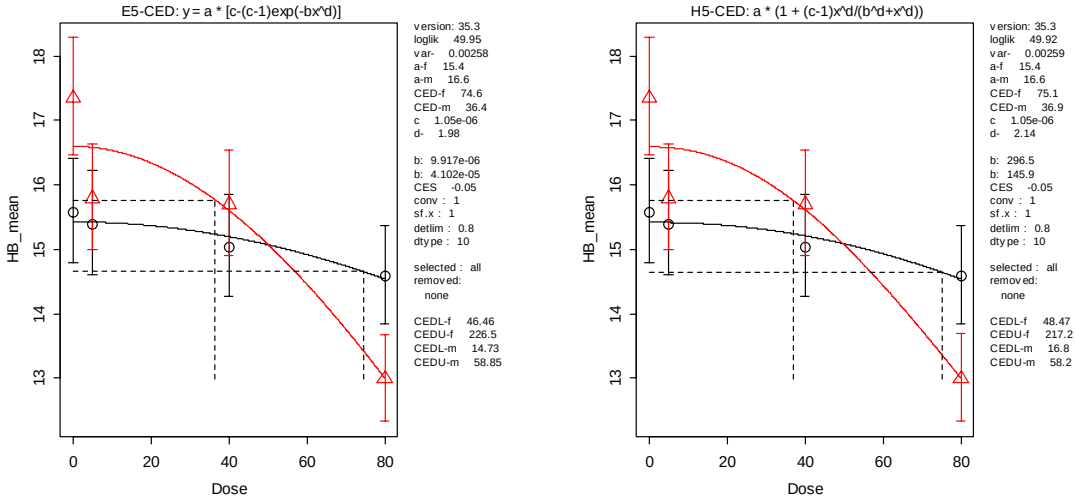
The usual models for dose-response analysis are the exponential and the Hill model. However, in PROAST the background response multiplies with the dose response (for various good reasons). In this particular case, where the background observations are not adequately reported, these models are not suitable. Therefore, we need to use a model that adds the background response to the dose-response, which is the case for the power models in PROAST.

A second (but related) reason to use a power model is the following. Normally, the BMR is defined as a percent change relative to background. Since the background response is not properly reported this measure does not work here. But we can use an absolute effect size, for instance a level of 2 (observation units). In PROAST this option exists for the power model.

Here, the females appear more sensitive (note, however, that an analysis with properly reported data and applying the usual models might give another result; i.e., background levels might already differ among the sexes). In this case the confidence interval is calculated by bootstrapping: (7.7, 46) units of MH

This is very close to the confidence interval for red blood cells. But note that the BMR of 2 units might be quite high, or more adverse than a 5% decrease in RBCs.

Decreased haemoglobin (one-year dog study)



BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
5% (increase)	14.729 – 58.846 (males) 46.456 – 226. 5 (females)	4 (males) 5 (females)

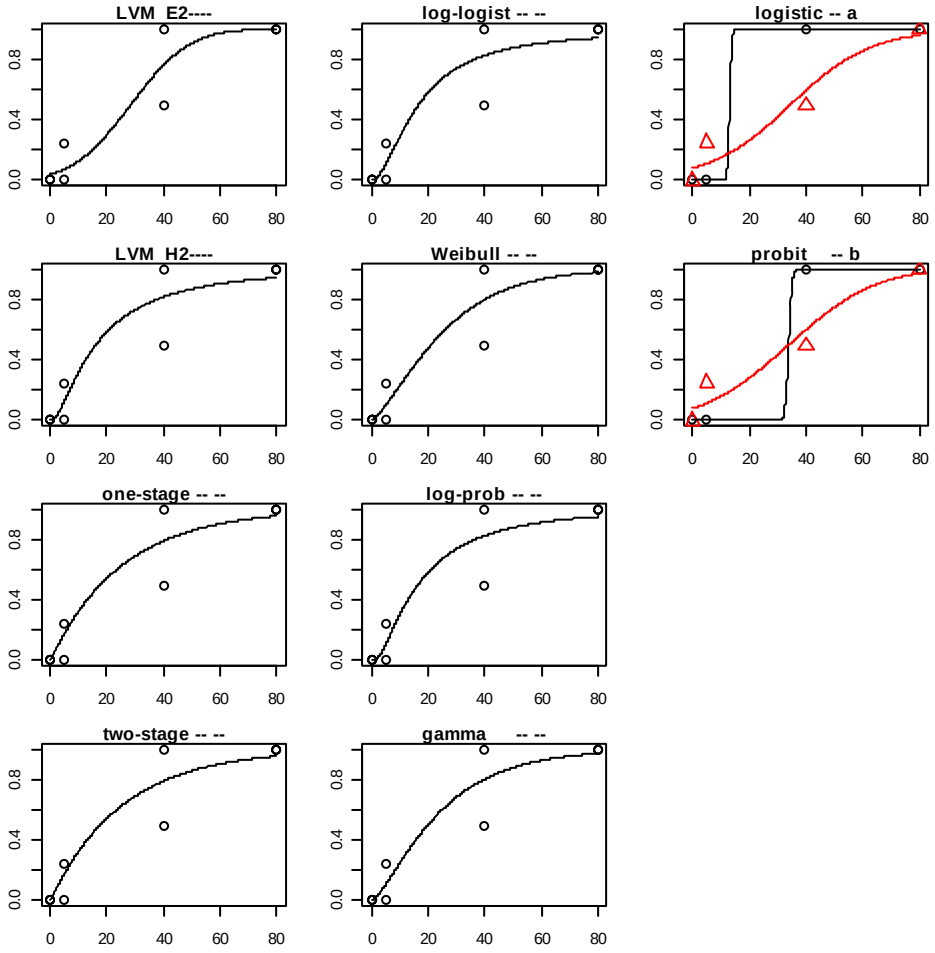
Comments

There was a significant sex difference, with males being more sensitive.

Increased liver hemosiderin (one-year dog study)

model	covar	npar	loglik	accept	BMD	BMDL	BMDU	level
null	NA	1	-22.12	--	NA	NA	NA	--
full	NA	8	-5.02	--	NA	NA	NA	--
one-stage	--	2	-7.99	yes	2.68	1.6	4.63	f

two-stage	--	3	-7.99	yes	2.68	1.6	4.63	f
log-logist	--	3	-8.15	yes	4.83	1.05	9.85	f
Weibull	--	3	-7.75	yes	4.96	0.689	22.1	f
log-prob	--	3	-8.06	yes	4.82	1.19	7.83	f
gamma	--	3	-7.8	yes	4.69	0.413	18.7	f
logistic	a	3	-5.89	yes	12.2	5.69	24.9	f
probit	b	3	-5.83	yes	11.1	NA	NA	m
LVM: E2-	--	2	-7.97	yes	10.6	5.95	18.5	-
LVM: H2-	--	2	-8.08	yes	4.39	1.9	7.64	-



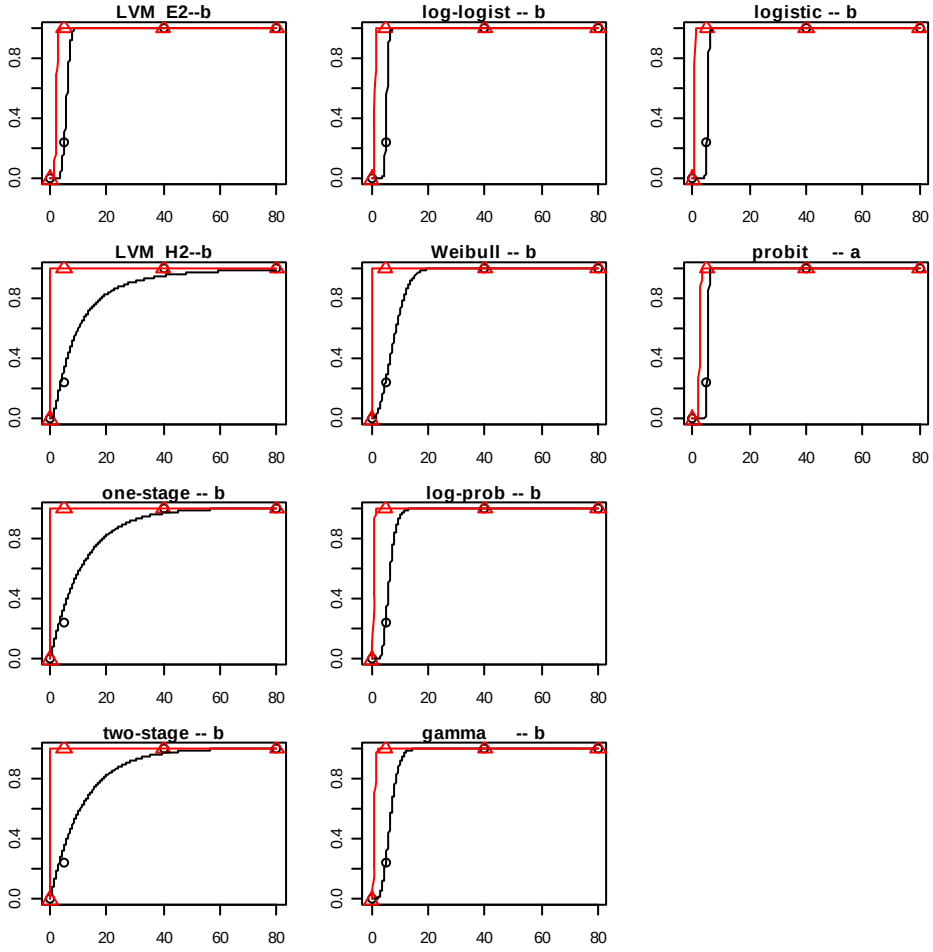
BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% extra risk	0.413 – 24.9	20

Comments
Significant sex difference, with females being more sensitive.

Increased kidney hemosiderin (one-year dog study)

model	covar	npar	loglik	accept	BMD	BMDL	BMDU	level
null	NA	1	-20.59	--	NA	NA	NA	--
full	NA	8	-2.25	--	NA	NA	NA	--
one-stage	b	3	-2.48	yes	0.000177	0	0.426	m
two-stage	b	4	-2.48	yes	4.59e-06	0	0.369	m

log-logist	b	4	-2.25	yes	1.05	0.646	1.19	m
Weibull	b	4	-2.25	yes	0.0257	0	0.106	m
log-prob	b	4	-2.25	yes	0.515	0	2.04	m
gamma	b	4	-2.25	yes	0.597	0	2.17	m
logistic	b	3	-2.25	yes	0.679	0.00898	5	m
probit	a	3	-2.25	yes	1.89	NA	NA	m
LVM: E2-	b	3	-2.25	yes	1.62	0	3.33	-
LVM: H2-	b	3	-2.56	yes	1e-06	0	0.756	-

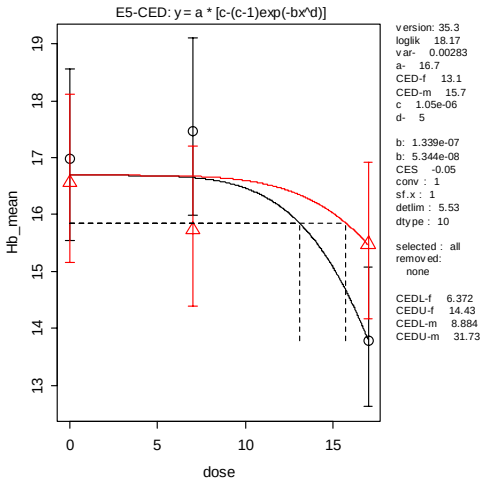


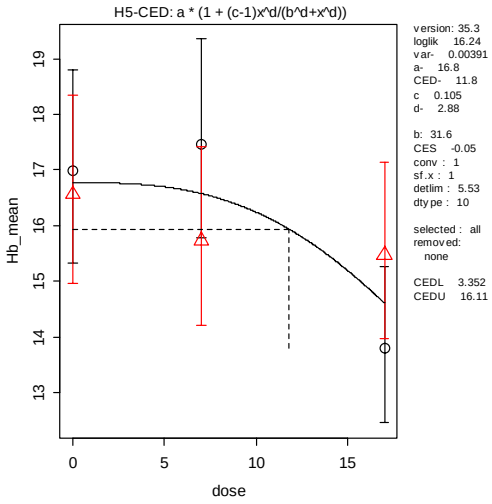
BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% extra risk	0 - 5	Infinite

Comments

There was a significant sex difference, with males being more sensitive. However, the data are not useful for a PoD because the lowest dose in males shows a nearly 100% response.

Decreased haemoglobin (30-day dog study)

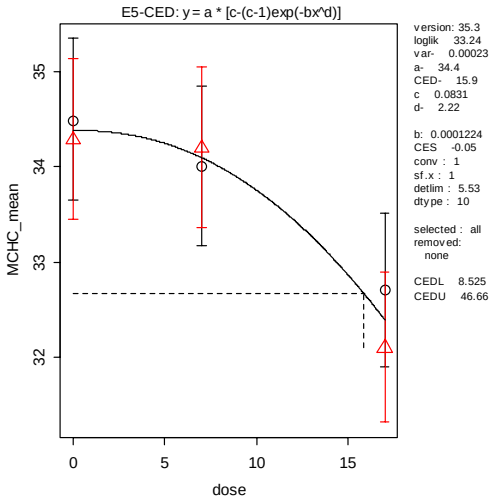


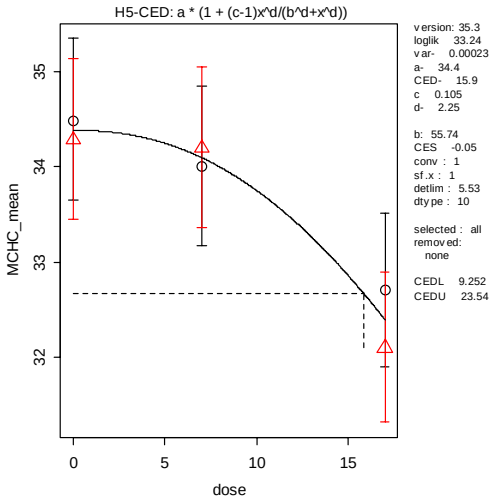


BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
5%	6.3 – 31.73	5

Comments
Sex difference is uncertain. Looking at these with effects on red blood cells, there is no consistent picture. The study is weak.

Decreased MCHC (30-day dog study)

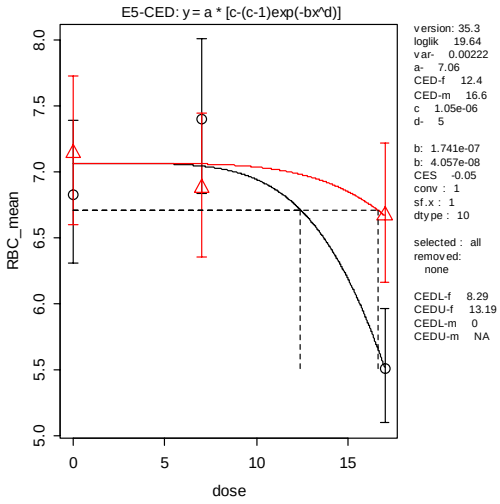


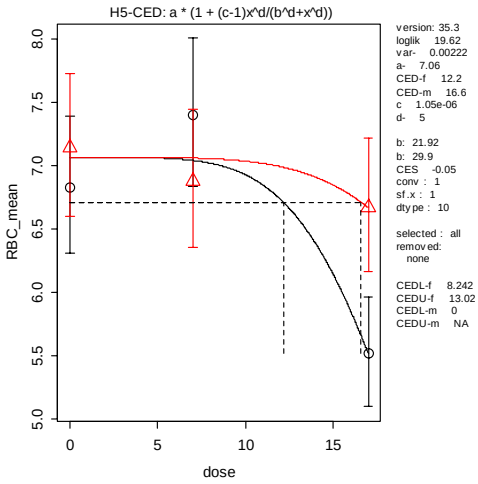


BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
5%	8.52 – 46.66	5

Comments
The data is okay. There were no significant sex differences.

Decreased red blood cells (30-day dog study)





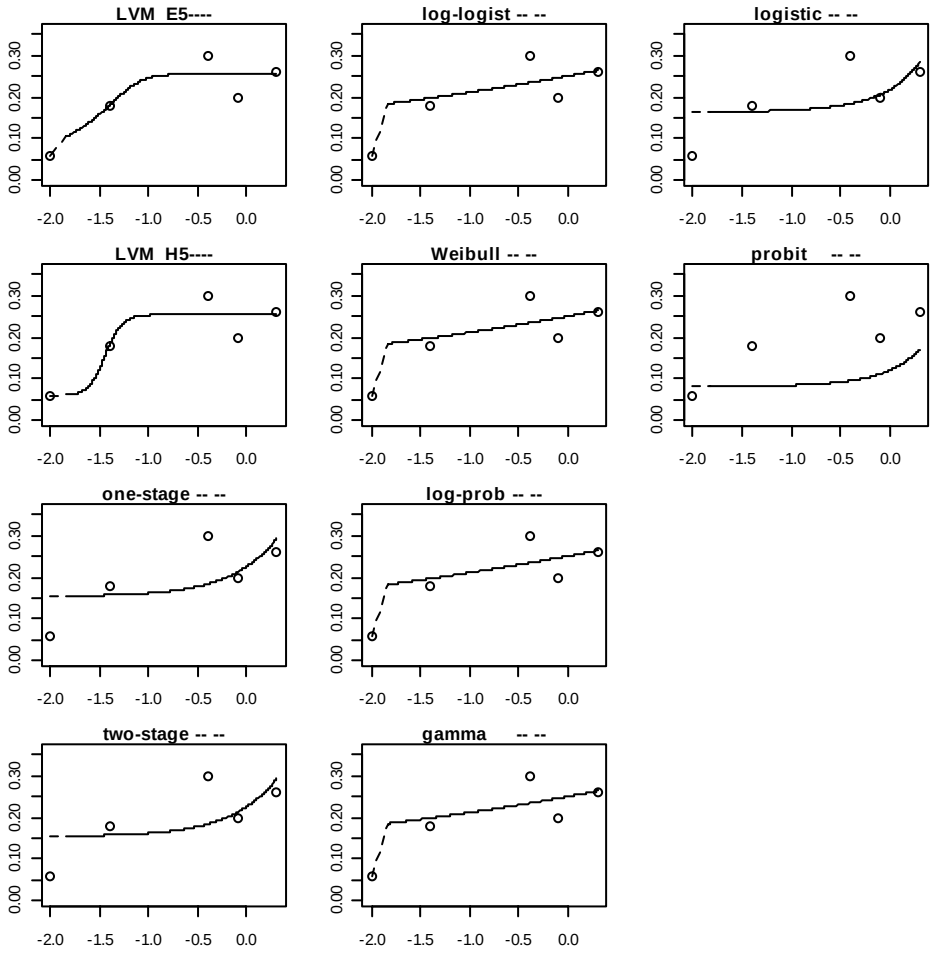
BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
5%	BMDL is 0 (males) 8.24 – 13.19 (females)	Infinite (males) 1.6 (females)

Comments

For males the BMDL is 0, so not useful data for PoD.

Sulcotrione

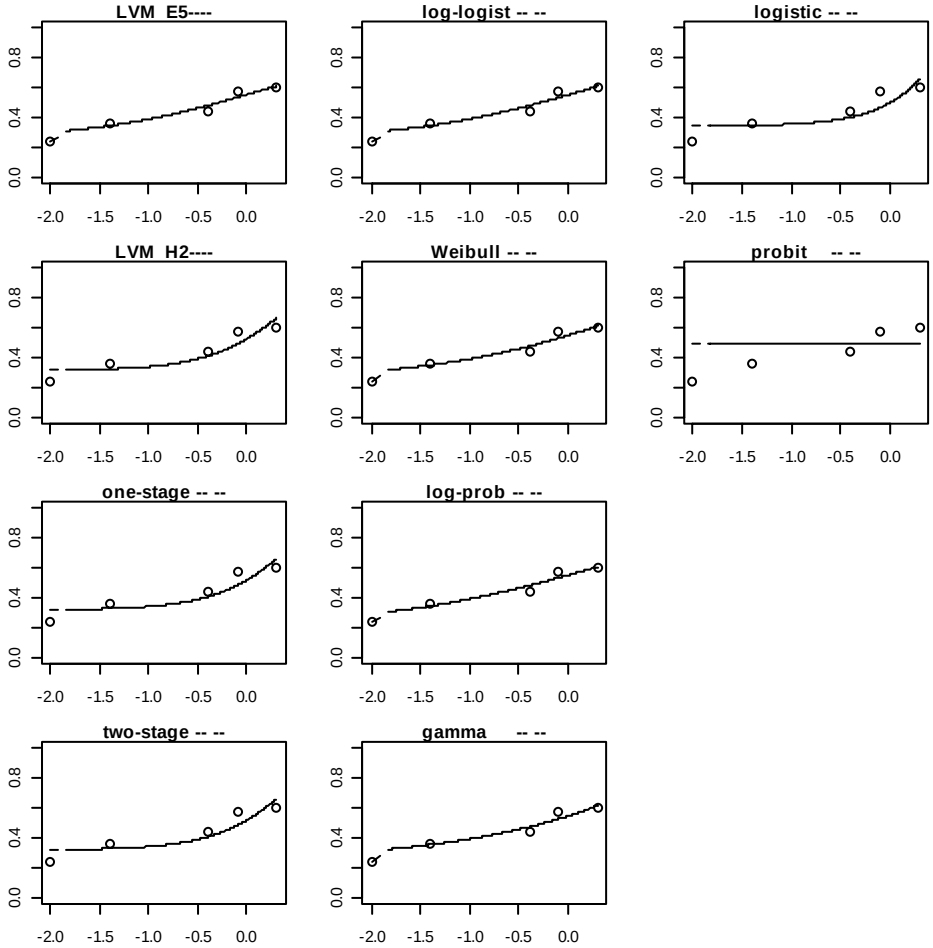
Kidney enlargement (two-year rat study)



BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% extra risk	0 – 0.39 (males)	Infinite

Comments
BMDL10 was 0 so PoD could not be derived.

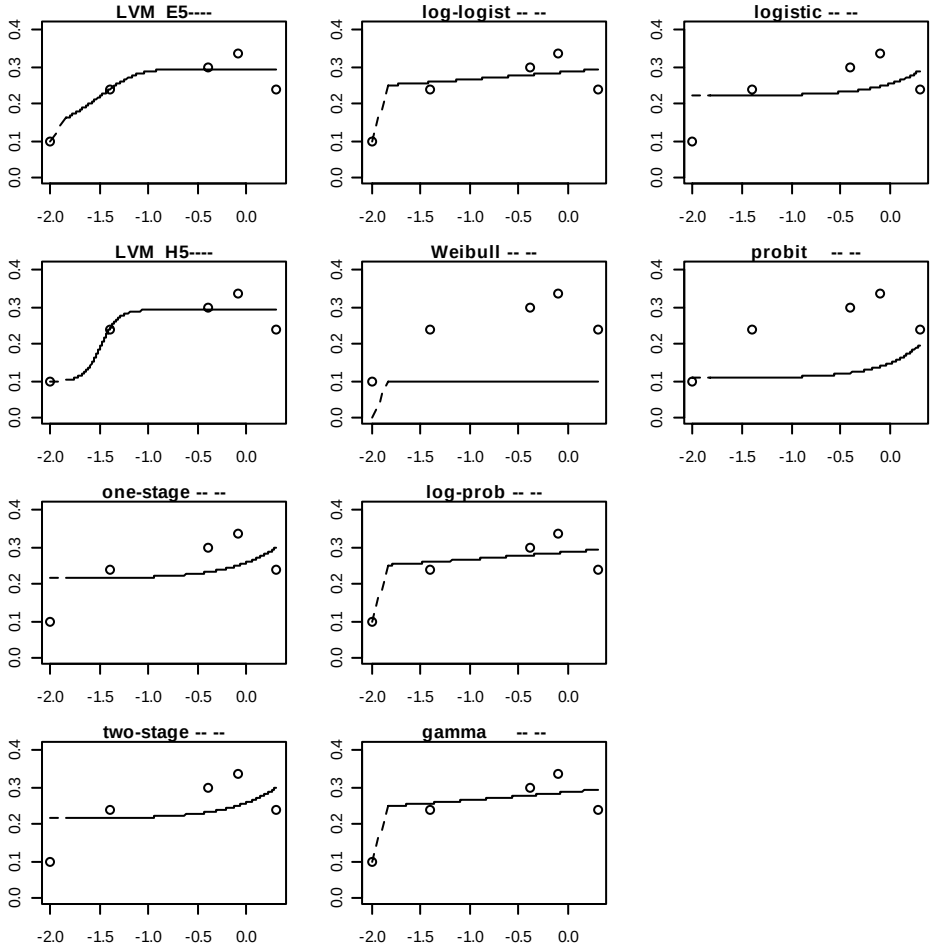
Kidney cystic change (two-year rat study)



BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% extra risk	0 – 0.74 (males)	Infinite

Comments
BMDL10 was 0 so PoD could not be derived.

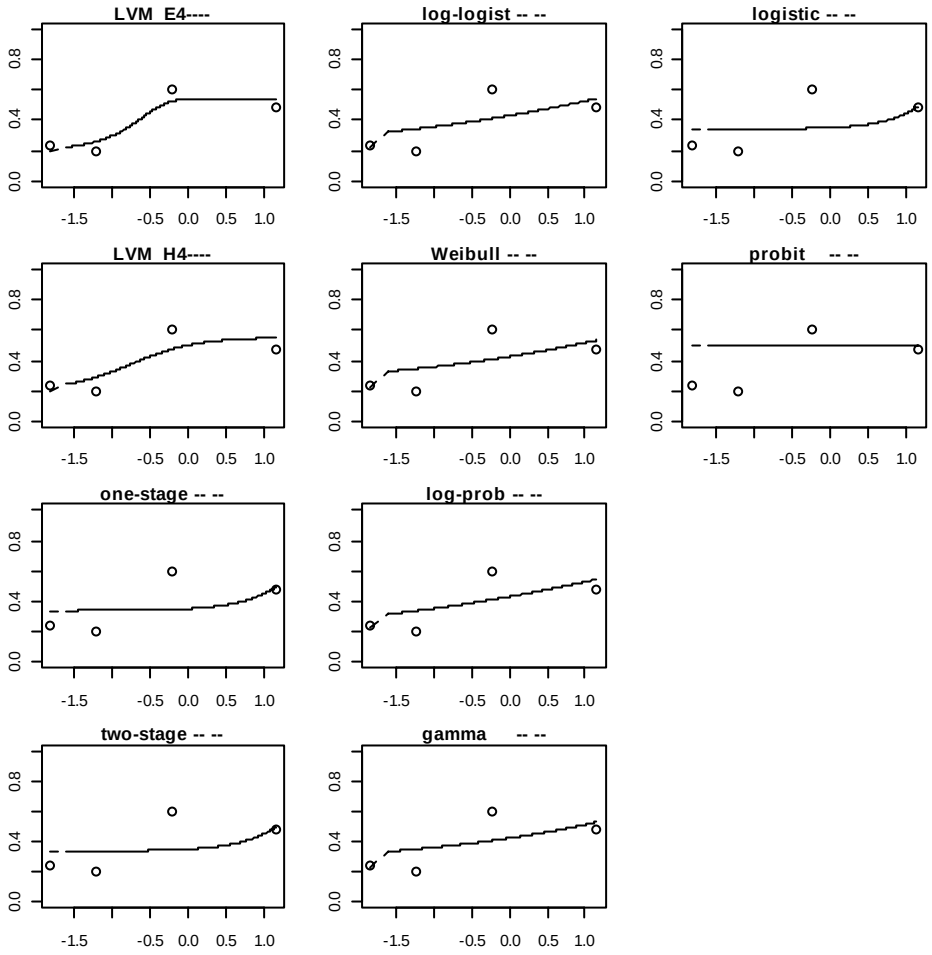
Kidney pelvic dilatation (two-year rat study)



BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% extra risk	0 – 0.415 (males)	Infinite

Comments
BMDL10 was 0 so PoD could not be derived.

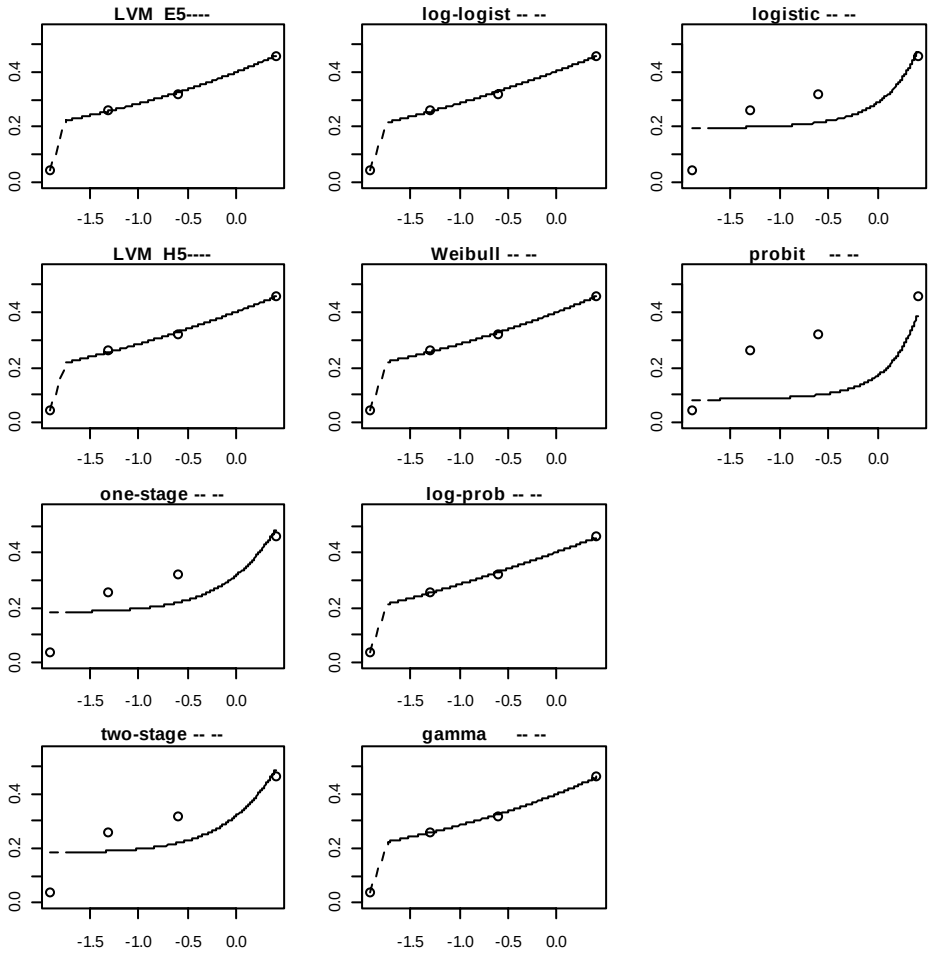
Nephropathy in P0 males (two-generation rat study)



BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% extra risk	0.0115 – 0.306	26

Triazoxide

Dark spleen (two-year rat study)



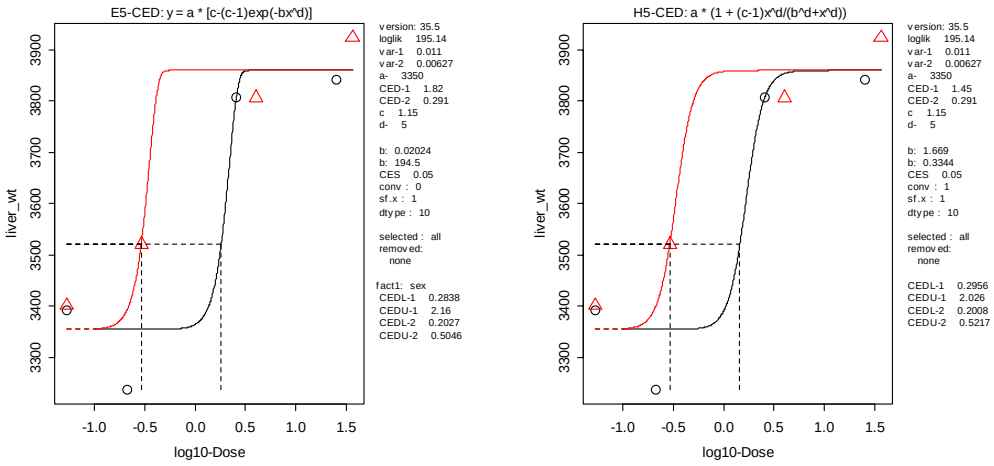
model	npar	loglik	accept	BMD	BMDL	BMDU
null	1	-116.65	--	NA	NA	NA
full	4	-102.89	--	NA	NA	NA
one-stage	2	-109.73	no	0.577	NA	NA
two-stage	3	-109.73	no	0.577	NA	NA
log-logist	3	-102.9	yes	0.0011	0	0.0256
Weibull	3	-102.9	yes	0.000686	0	0.0226
log-prob	3	-102.91	yes	0.00161	0	0.0287
gamma	3	-102.89	yes	0.00038	0	0.0196
logistic	2	-110.18	no	0.878	NA	NA
probit	2	-119.88	no	1.03	NA	NA
LVM: E5-	4	-102.89	yes	0.000324	0	0.0179
LVM: H5-	4	-102.9	yes	0.000876	0	0.0236

BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*

10% extra risk	0 – 0.0287	Infinite

Comments
There is a significant trend, but the lower bound for the BMD10 is "zero". In the NOAEL approach the lowest dose is a LOAEL. It is not possible to derive a PoD from these data. The study should have had more (lower) doses.

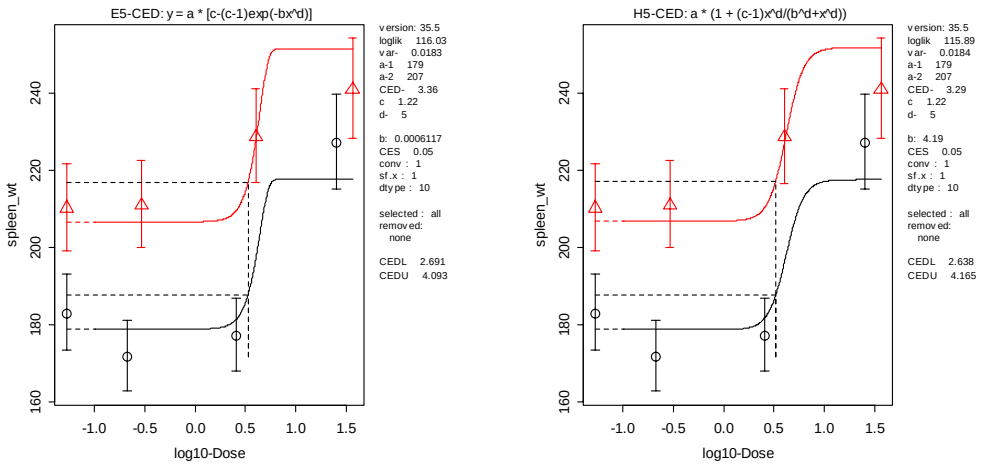
Increased liver weight (90-day rat study)



BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
5% (increase)	0.28 – 2.16 (males)	7.7 (males)
	0.20 – 0.52 (females)	2.6 (females)

Comments
Females were more sensitive. The imprecision is good.

Increased spleen weight (90-day rat study)

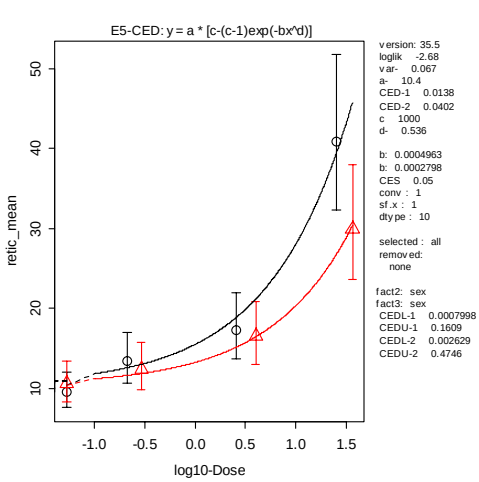


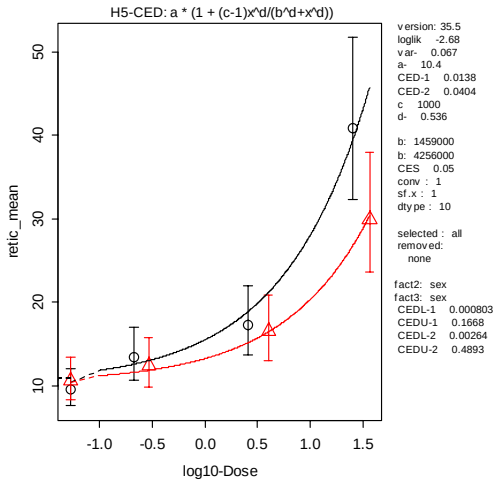
BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*

5% (increase)	2.64 – 4.16	1.6

Comments
No sex difference. Very good imprecision, indicating good data.

Increased reticulocytes (90-day rat study)





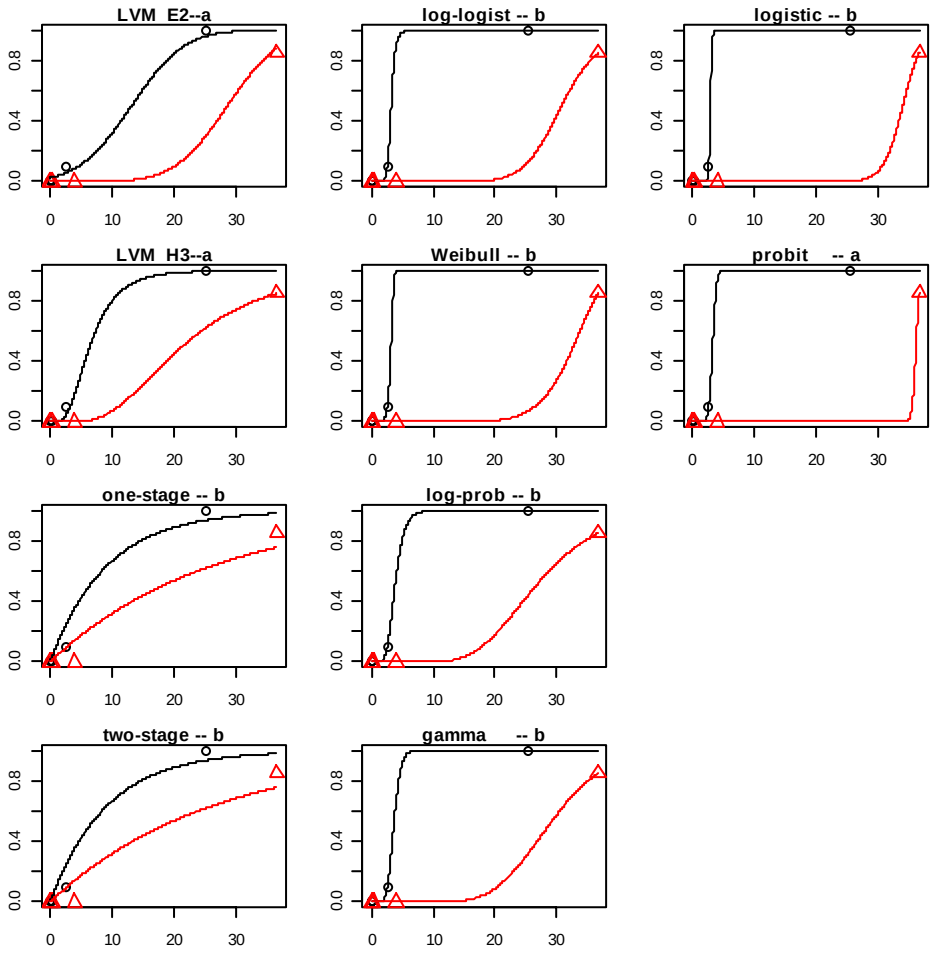
BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
5% (increase)	8×10^{-4} – 0.167 (males)	Very large
	0.0026 – 0.489 (females)	Very large
20% (increase)	0.027 – 0.881 (males)	33
	0.0879 – 2.40 (females)	27
25% (increase)	0.0477 – 1.185 (males)	25
	0.1537 – 2.401 (females)	16
50% (increase)	0.271 – 3.009 (males)	11
	0.856 – 8.102 (females)	9

Comments
The BMD confidence intervals were very wide with a BMR of 5%. However, the maximum response is very large, around a 4-fold increase compared with controls. Hence, the higher BMRs gave a better imprecision factor.

Spleen congestion

model	covar	npar	loglik	accept	BMD	BMDL	BMDU	level
null	NA	1	-87.16	--	NA	NA	NA	--
full	NA	8	-14.85	--	NA	NA	NA	--
one-stage	b	3	-21.69	no	0.959	NA	NA	1
two-stage	b	4	-21.69	no	0.959	NA	NA	1
log-logist	b	4	-14.85	yes	2.56	1.8	3.68	1
Weibull	b	4	-14.85	yes	2.55	1.67	4.32	1
log-prob	b	4	-14.85	yes	2.55	1.77	3.56	1
gamma	b	4	-14.85	yes	2.55	1.7	3.92	1
logistic	b	3	-14.85	yes	2.56	2.04	3.81	1

probit	a	3	-14.85	yes	2.56	NA	NA	1
LVM: E2-	a	3	-16.93	yes	20.1	3.61	7.94	-
LVM: H3-	a	4	-14.98	yes	11	1.67	4.28	-

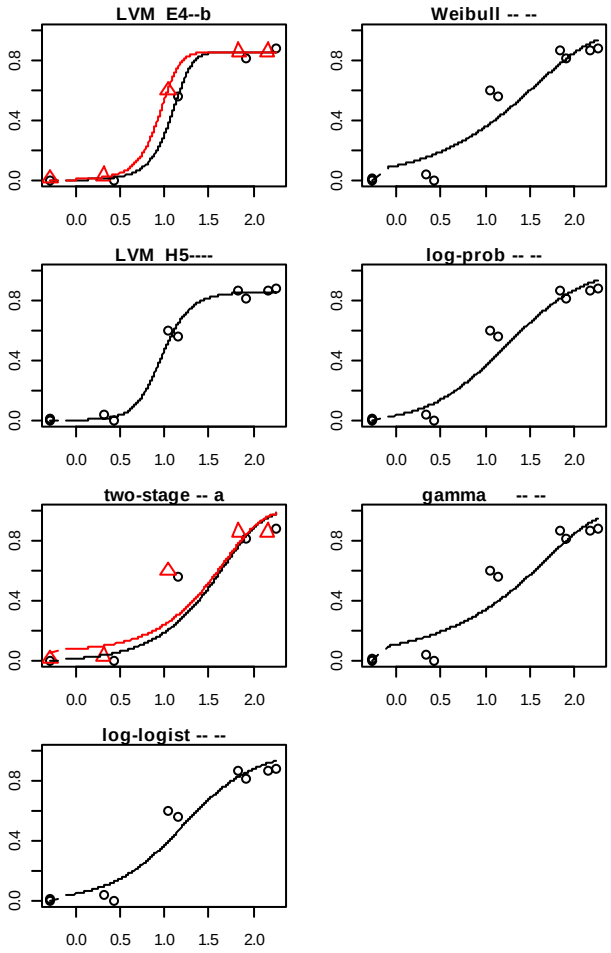


BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% extra risk	1.67 – 7.94 (males)	5 (males)

Comments
Effect in males only.

Fluapyroxad

Liver hypertrophy (two-year rat study)



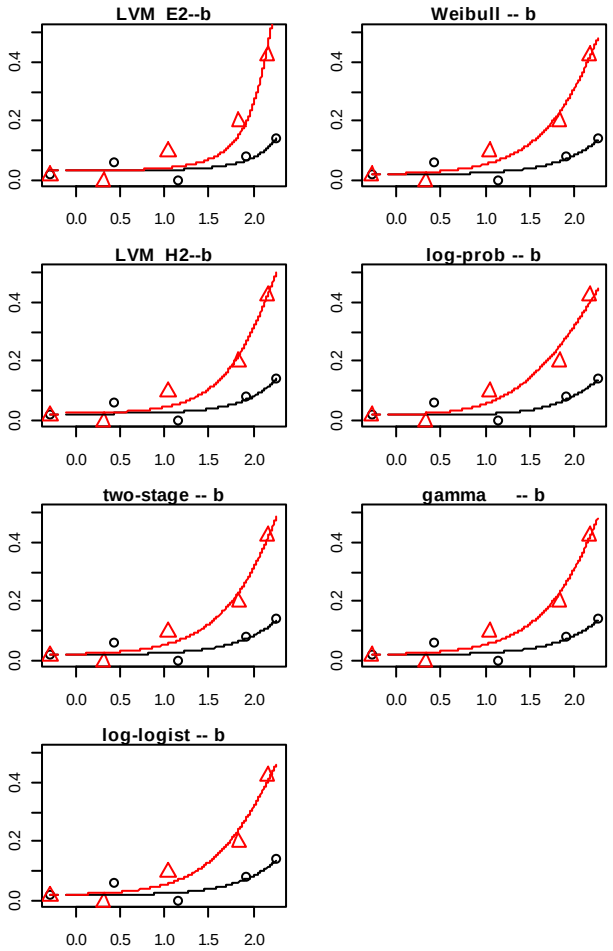
model	covar	npar	loglik	accept	BMD	BMDL	BMDU	level
null	NA	1	-345.28	--	NA	NA	NA	--
full	NA	10	-163.66	--	NA	NA	NA	--
two-stage	a	4	-202.68	no	4.95	NA	NA	f
log-logist	--	3	-180.01	no	2.14	NA	NA	f
Weibull	--	3	-187.72	no	1.05	NA	NA	f
log-prob	--	3	-180.07	no	2.32	NA	NA	f
gamma	--	3	-190.76	no	0.872	NA	NA	f
LVM: E4-	b	4	-166.29	yes	4.42	3.47	5.45	-
LVM: H5-	--	4	-167.91	yes	4.72	3.33	7.01	-

BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% extra risk	3.33 - 7.01	2

Comments

Only two models were accepted, normally you would decrease the P-value but in this case the other models will not be accepted at even very low P-values. NOTE: these two latent variable model are not included in BMDS.

Liver tumours (two-year rat study)



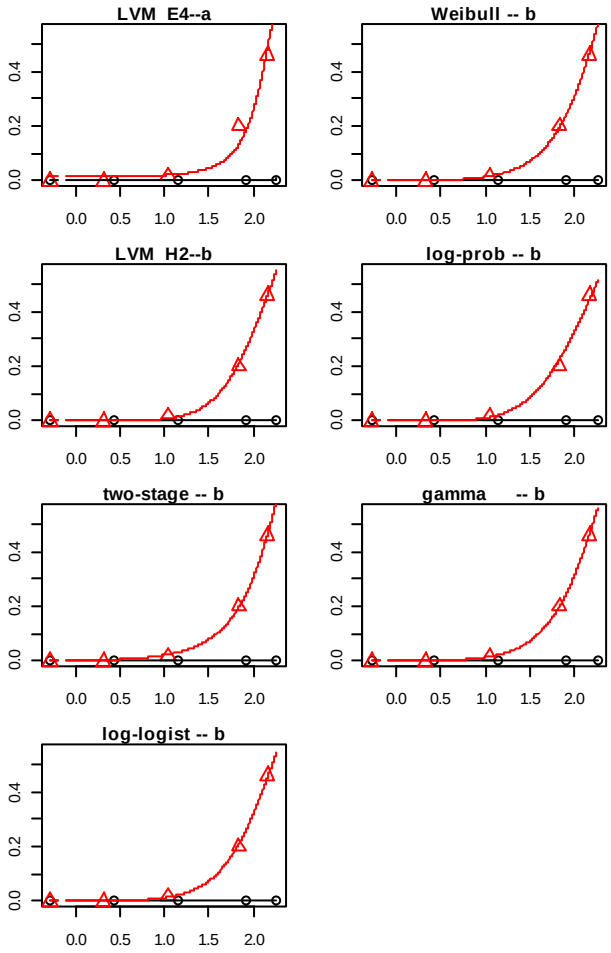
model	covar	npar	loglik	accept	BMD	BMDL	BMDU	level
null	NA	1	-166.89	--	NA	NA	NA	--
full	NA	10	-130.63	--	NA	NA	NA	--
two-stage	b	4	-135.48	yes	30.1	22.5	42.2	m
log-logist	b	4	-135.51	yes	28.1	12.9	54.4	m
Weibull	b	4	-135.48	yes	29.3	13.3	55.6	m
log-prob	b	4	-135.39	yes	24.5	12.3	46.9	m
gamma	b	4	-135.48	yes	29.4	13.5	54.3	m
LVM: E2-	b	3	-137.04	yes	59.8	50.6	71.4	-
LVM: H2-	b	3	-135.73	yes	36.6	27.9	49.3	-

BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% extra risk	12.3 – 71.4	6

Comments

Significant sex difference, with males (red) being more sensitive.

Liver spongiosis (two-year rat study)



model	covar	npar	loglik	accept	BMD	BMDL	BMDU	level
null	NA	1	-124.22	--	NA	NA	NA	--
full	NA	10	-64.42	--	NA	NA	NA	--
two-stage	b	4	-64.69	yes	28.3	23.4	59.2	m
log-logist	b	4	-64.52	yes	38.9	23.4	55.7	m
Weibull	b	4	-64.52	yes	39.8	23.8	57	m
log-prob	b	4	-64.63	yes	34.7	21.1	49.8	m
gamma	b	4	-64.51	yes	39.2	23.9	55.2	m
LVM: E4-	a	4	-67.37	yes	62.2	NA	NA	-
LVM: H2-	b	3	-64.75	yes	40.7	32.1	51.5	-

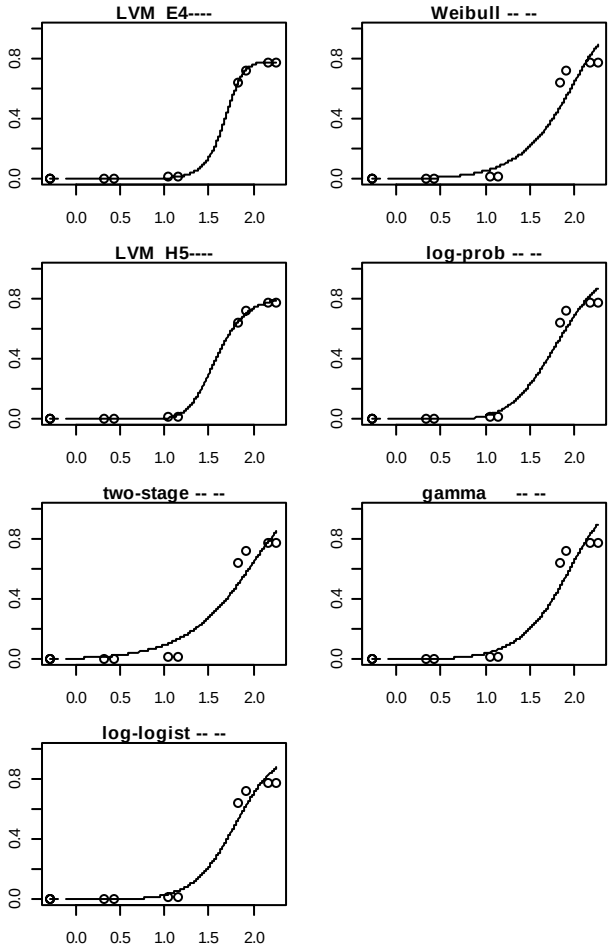
BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% extra risk	21.1 – 59.2	3

Comments

Significant sex difference, with males (red) being more sensitive.

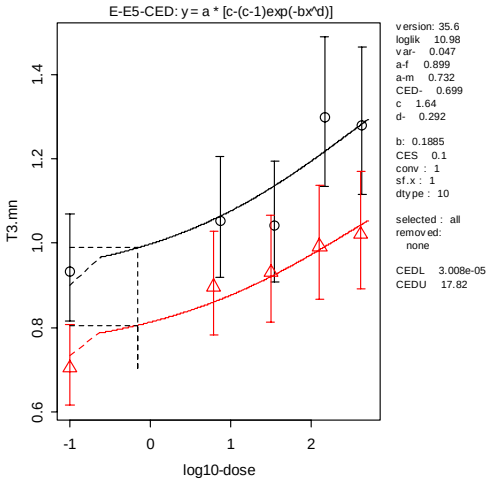
Tooth whitening

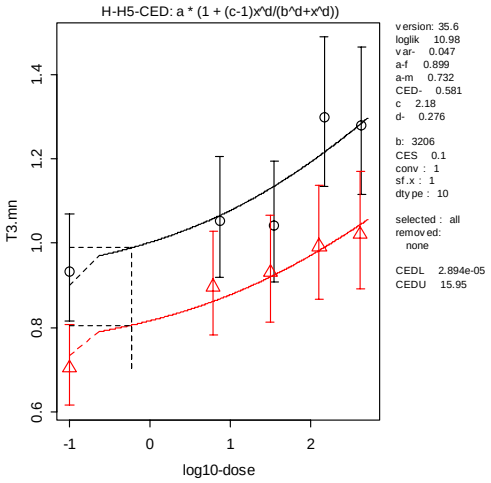
model	covar	npar	loglik	accept	BMD	BMDL	BMDU	level
null	NA	1	-303.72	--	NA	NA	NA	--
full	NA	10	-124.81	--	NA	NA	NA	--
two-stage	--	3	-139.4	no	10.2	NA	NA	f
log-logist	--	3	-130.57	yes	19.1	14.2	24.9	f
Weibull	--	3	-137	no	15.8	NA	NA	f
log-prob	--	3	-129.94	yes	18.7	14.1	23.9	f
gamma	--	3	-135.42	no	18.2	NA	NA	f
LVM: E4-	--	3	-125.35	yes	27.8	18.5	34.5	-
LVM: H5-	--	4	-125.14	yes	19.9	15.6	29.8	-



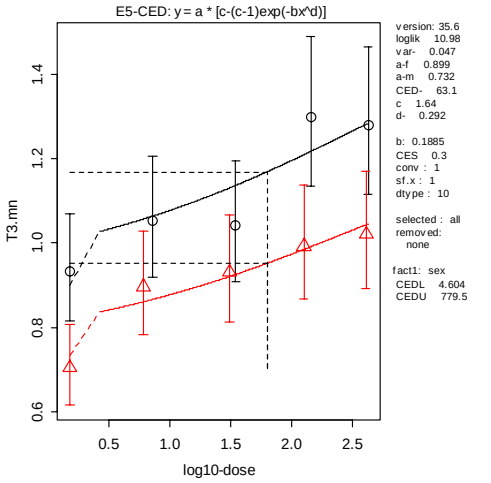
BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% extra risk	14.1 – 23.9	1.7

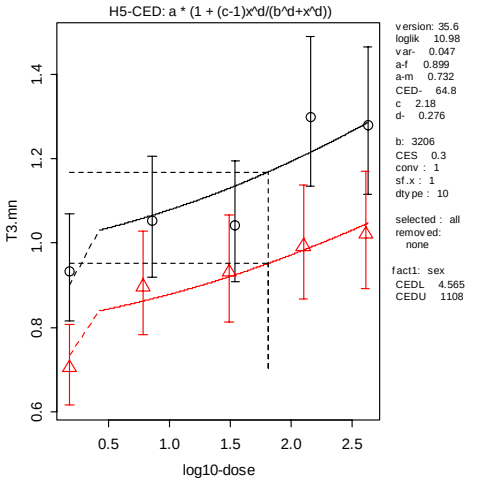
Increased T3 (90-day rat study)





BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% (increase)	BMDL = 0	Infinite



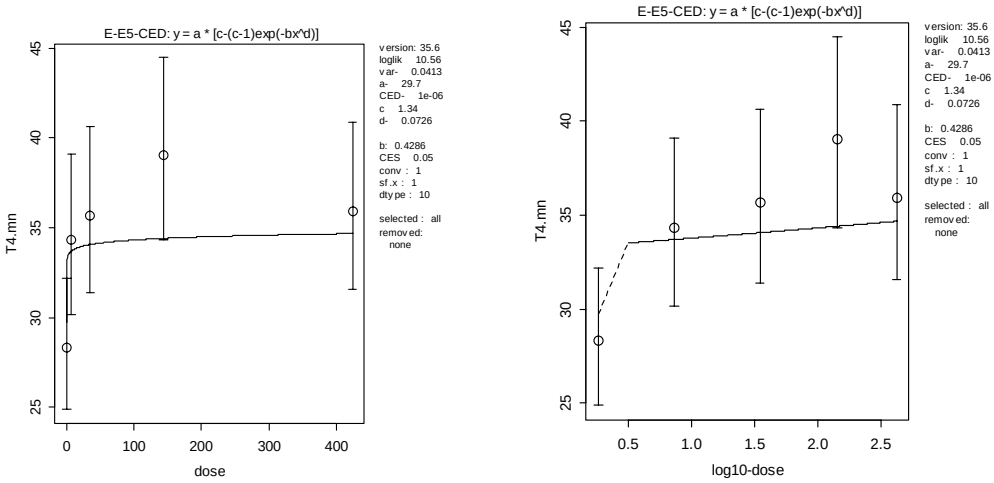


BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
30%	4.565 - 1107.5	246

Comments

There is a non-zero BMDL for BMR of 10%. The size of the effect at the lowest dose is much higher than 10%; in other words, the BMD10 is much lower than the lowest tested dose. Choosing a larger value of BMR (30%) results in a non-zero value for the BMDL, but the confidence interval is still large.

Increased T4 (90-day rat study)

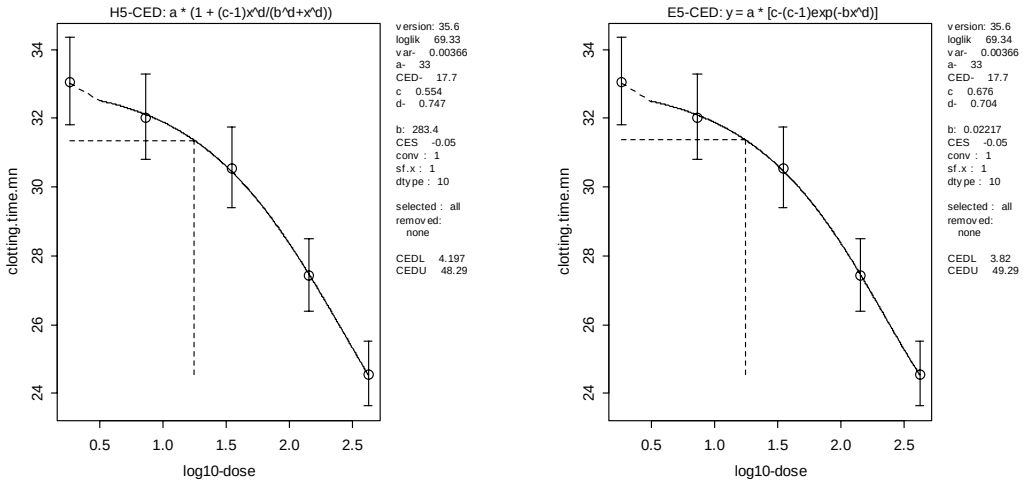


BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
5% (increase)	Infinitely large	Infinite

Comments

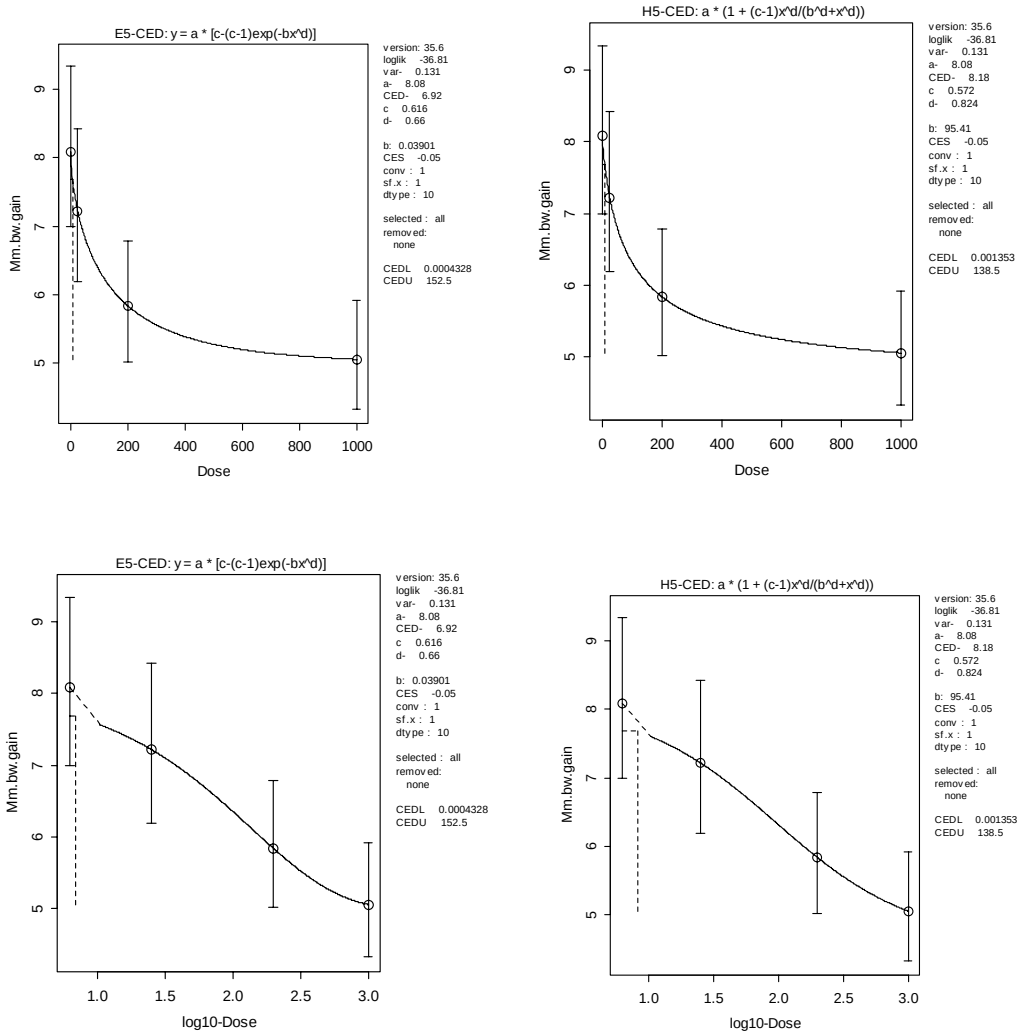
The confidence intervals for BMD are infinitely large. These data are problematic. The non-zero dose groups show responses that are not significantly increasing. It seems that the controls deviate from the other dose groups for other reasons than the dose. We know from other studies that a dose group can be an outlier in toxicology studies, and it seems that this occurs more often in the controls. Another indication is the small value of parameter c (1.3), which means that the response would level off at a 30% increase. In other data the maximum response is much larger for T4.

Decreased clotting time (90-day rat study)



BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
5% (decrease)	3.82 – 49.29	13
10% (decrease)	24.24 – 96.89	4

Decreased maternal body weight (rat developmental study)



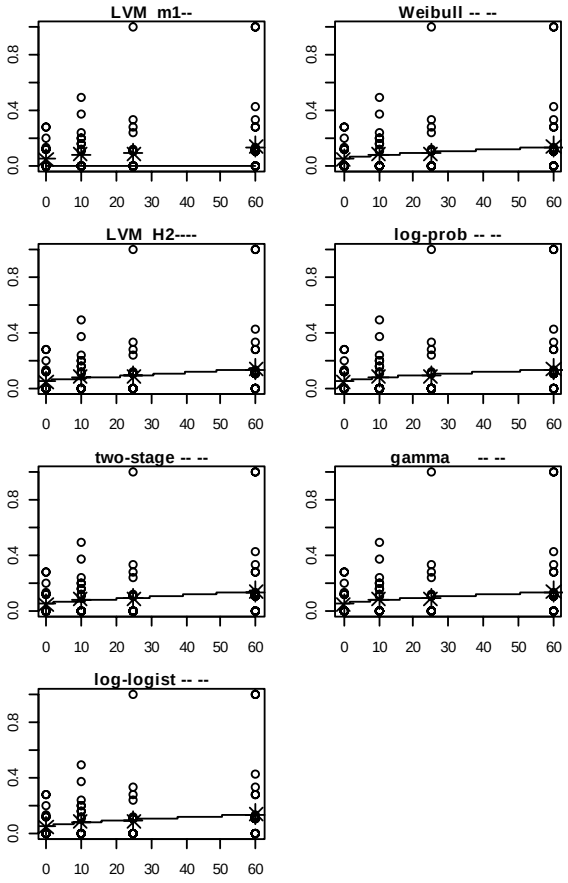
BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
5% (decrease)	Very large	Very large
15% (decrease)	0.767 – 224.2	300

Comments

A 5% decrease in body weight gain is clearly far beyond the range of observation (see lower plots on log-dose scale), hence the extremely large confidence intervals. Increasing the BMR to 15% results still gives a large imprecision factor (300). The reason is the large within-group variance.

Note that BW gain, measured as a difference in BW at the end and start of the study, is not an appropriate measure. Body weight gain should be measured as the ratio of BWs. Or, just BW at the end of the study also reflects changes in growth, since there is a control group.

Resorptions (rabbit developmental study)



model	npar	loglik	accept	BMD	BMDL	BMDU
null	2	-185.76	--	NA	NA	NA
full	5	-183.67	(no)	NA	NA	NA
two-stage	3	-183.79	yes	66.9	32.6	413
log-logist	3	-183.73	yes	73.7	25.7	3e+09
Weibull	3	-183.73	yes	73.1	25.8	2.96e+09
log-prob	3	-183.74	yes	76.8	25.3	4460000
gamma	3	-183.73	yes	72.8	26	2.94e+09
LVM: m1-	2	-183.78	yes	NA	NA	NA
LVM: H2-	3	-183.81	yes	66.5	34.3	395

BMR	*BMD confidence interval*	*Imprecision factor (BMDU/BMDL)*
10% extra risk	25.3 - infinite	Infinite

Comments

The overall BMD confidence interval is 25.3 to infinite is infinite (in practical terms). So, it is quite likely that the effect size of 10% extra risk will never be reached, at least not for practically feasible doses. Note that there are only two nests with 100% response, and they happen to be at the two highest doses. Had the animal at the highest dose not been included in the study, the response would have been statistically non-significant.

ANNEX 5

Strategies for literature searches and databases searched.

WinSPIRS 5.00

No.	Records	Request
1	88769	chemical
2	33269	chemicals
3	106893	chemical or chemicals
4	101067	method*
5	42812	technique*
6	629	paradigm*
7	98270	determin*
8	183881	method* or technique* or paradigm* or determin*
9	71883	risk
10	43794	assessment*
11	16688	risk near3 assessment*
12	15752	reference
13	53250	dose*
14	15752	reference
15	43416	value*
16	2074	allowable
17	13415	daily
18	4782	intake*
19	4621	acceptable
20	13415	daily
21	4782	intake*
22	2074	allowable
23	5441	operator
24	159113	exposure
25	84189	level*
26	4621	acceptable
27	5441	operator
28	159113	exposure
29	84189	level*
30	673	(reference adj dose*) or (reference adj value*) or (allowable adj daily adj intake*) or (acceptable adj daily adj intake*) or (allowable adj operator adj exposure adj level*) or (acceptable adj operator adj exposure adj level*)
31	323	benchmark
32	53250	dose*
33	40171	workplace
34	159113	exposure
35	84189	level*
36	9113	compound
37	31215	specific
38	3085	adjustment
39	90253	factor*
40	9113	compound
41	31215	specific
42	41198	assessment
43	90253	factor*
44	88769	chemical
45	31215	specific
46	3085	adjustment
47	90253	factor*
48	105	(benchmark adj dose*) or (workplace adj exposure adj level*) or (compound adj specific adj adjustment adj factor*) or (compound adj specific adj assessment adj factor*) or (chemical adj specific adj adjustment adj factor*)
49	88769	chemical
50	31215	specific
51	41198	assessment
52	90253	factor*
53	67	ADI
54	53	BMD
55	0	AOEL
56	14	WEL
57	0	CSAF
58	134	(chemical adj specific adj assessment adj factor*) or ADI or BMD or AOEL or WEL or CSAF
59	842	#30 or #48 or #58
60	19	#3 and #8 and #11 and #59
61	8270	PY = 2007-2012
62	1	#60 and (PY = 2007-2012)
63	142	#3 and #8 and #59
64	8270	PY = 2007-2012
65	3	#63 and (PY = 2007-2012)
66	255	#3 and #59
67	8270	PY = 2007-2012
68	6	#66 and (PY = 2007-2012)
69	471	#8 and #59

```
? d s
Set     Items   Description
S1    2123135   S CHEMICAL OR CHEMICALS
S2   11103620   S METHOD? OR TECHNIQUE? OR PARADIGM? OR DETERMINE?
S3     528794   S RISK (3N) (ASSESSMENT? OR ESTIMATE?)
S4     215344   S (REFERENCE (W) DOSE?) OR (REFERENCE (W) VALUE?) OR (ALLOWABLE (W) DAILY
(W) INTAKE?) OR (ACCEPTABLE (W) DAILY (W) INTAKE?) OR (ALLOWABLE (W) OPERATOR (W)
EXPOSURE (W) LEVEL?) OR (ACCEPTABLE (W) OPERATOR (W) EXPOSURE (W) LEVEL?) OR (BENCHMARK
(W) DOSE?) OR (WORKPLACE (W) EXPOSURE (W) LEVEL?) OR (COMPOUND (W) SPECIFIC (W)
ADJUSTMENT (W) FACTOR?) OR (COMPOUND (W) SPECIFIC (W) ASSESSMENT (W) FACTOR?) OR
(CHEMICAL (W) SPECIFIC (W) ADJUSTMENT (W) FACTOR?) OR (CHEMICAL (W) SPECIFIC (W)
ASSESSMENT (W) FACTOR?) OR ADI OR BMD OR AOEL OR WEL OR CSAF
S5   11574461   S METHOD? OR TECHNIQUE? OR PARADIGM? OR DETERMIN?
S6     533744   S RISK (3N) (ASSESSMENT? OR ESTIMAT?)
S7        830   S S1 AND S5 AND S6 AND S4
S8        279   S S7/2007:2012
S9        262   S S8/ENG
S10       134   RD  (unique items)
S11     36896   S S1 AND S6
S12    165397   S S4/TI,DE
S13       589   S S11 AND S12
S14       159   S S13/2007:2012
S15       150   S S14/ENG
S16        92   RD  (unique items)
S17        37   S S16 NOT S10
S18    718606   S S1 AND S5
S19      3306   S S18 AND S12
S20       840   S S19/2007:2012
S21       794   S S20/ENG
S22       614   RD  (unique items)
S23       559   S S22 NOT S10
S24       554   S S23 NOT S17
```

WinSPIRS 5.00

No.	Records	Request
1	88769	chemical
2	33269	chemicals
3	106693	chemical or chemicals
4	22505	approach*
5	101067	method*
6	42812	technique*
7	629	paradigm*
8	98270	determin*
9	15570	deriv*
10	203595	approach* or method* or technique* or paradigm* or determin* or deriv*
11	44614	#3 and #10
12	15752	reference
13	43416	value*
14	53250	dose*
15	553	reference adj (value* or dose*)
16	159113	exposure
17	84189	level*
18	43416	value*
19	16846	exposure near3 (level* or value*)
20	71883	risk
21	69572	assess*
22	17639	risk near3 assess*
23	2074	allowable
24	4621	acceptable
25	13415	daily
26	4782	intake*
27	128	(allowable or acceptable) adj daily adj intake*
28	4621	acceptable
29	2074	allowable
30	159113	exposure
31	84189	level*
32	26	(acceptable or allowable) adj exposure adj level*
33	323	benchmark
34	53250	dose*
35	80	benchmark adj dose*
36	902	physiologically
37	41482	based
38	1163	pharmacokinetic
39	38386	model*
40	238	physiologically adj based adj pharmacokinetic adj model*
41	40171	workplace
42	159113	exposure
43	84189	level*
44	54082	limit*
45	83	workplace adj exposure adj (level* or limit*)
46	8089	derived
47	213453	no
48	69632	effect
49	84189	level*
50	2	derived adj no adj effect adj level*
51	213453	no
52	38927	observed
53	7880	adverse
54	69632	effect
55	84189	level*
56	117	no adj observed adj adverse adj effect adj level*
57	88769	chemical
58	9113	compound
59	31215	specific
60	3085	adjustment
61	41198	assessment
62	90253	factor*
63	0	(chemical or compound) adj specific adj (adjustment or assessment) adj factor*
64	0	aoel
65	47	ard
66	67	adi
67	53	bmd
68	319	pbpk
69	0	dnel
70	230	noael
71	0	csaf
72	698	aoel or ard or adi or bmd or pbpk or dnel or noael or csaf
73	34926	#15 or #19 or #22 or #27 or #32 or #40 or #45 or #50 or #56 or #63 or #72
74	4334	#11 and #73
75	85711	PY > "2000"
76	1017	#74 and (PY > "2000")

Databases searched were:

OSHROM (comprising HSEline, Cisdoc, Nioshtic, Rilosh, Oshline)

Web of Knowledge(Comprising Web of Science and Medline)

Chemical abstracts, Chemical Business Newsbase, Chemical Safety Newsbase, Embase, Toxfile, Pharmaceutical News Index and Healsafe.

70	8270	PY = 2007-2012
71	17	#69 and (PY = 2007-2012)
72	88	#11 and #59
73	8270	PY = 2007-2012
74	4	#72 and (PY = 2007-2012)