For Official Use **ENV/JM/TG/M(2008)5/REV**

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development **19-Jan-2009**

English - Or. English

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND
THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

ENV/JM/TG/M(2008)5/REV
For Official Use

Test Guidelines Programme

SUMMARY RECORD OF THE EXPERT GROUP MEETING ON THE EXTENDED ONE GENERATION REPRODUCTIVE TOXICITY STUDY

15-17 October 2008, Paris Headquarters, France

Nathalie DELRUE
Tel: +33 (0)1 45 24 98 44; Fax: +33 (0)1 44 30 61 80; Email: nathalie.delrue@oecd.org

JT03258335

English - Or. English

SUMMARY RECORD OF THE EXPERT GROUP MEETING ON THE EXTENDED ONE GENERATION REPRODUCTIVE TOXICITY STUDY

(15-17 OCTOBER 2008, PARIS, FRANCE)

Opening of the meeting and approval of the draft agenda

1. Aldert Piersma (NL) and Elizabeth Mendez (US) co-chaired the meeting in turn. The Chair welcomed the participants and invited them to introduce themselves. Experts from Canada, Denmark, Germany, Japan, Netherlands, Sweden, Switzerland, United Kingdom, United States, European Commission (EC/ECVAM and EC/ECHA), BIAC, and ICAPO attended the meeting.

2. The list of participants is available at the Secretariat for government representatives.

3. The group adopted the draft agenda [ENV/JM/TG/A(2008)5]. ICAPO and BIAC noted that a significant policy change was introduced into version 5 of the draft TG (i.e. changing Cohorts 2 and 3 from optional to mandatory components of the study) and questioned the appropriateness of using version 5 as a starting point without the expert group's consideration and agreement on this point. The Chair indicated that this discussion would take place during the plenary discussion.

Introduction by the Secretariat

4. The Secretariat presented the documents for the meeting, provided the history since the development of the project started at OECD (led by US, Germany and NL) and reminded the participants of the objective of the meeting: (i) agree on a study design and (ii) address the technical issues raised by the comments on the draft TG. The Secretariat also indicated that following the meeting, the draft TG would be circulated to the WNT for a second commenting round in November 2008. If the group agrees on a study design at this meeting, then the draft TG could be submitted for approval at the next WNT meeting in April 2009, provided the comments received from the WNT during the second commenting round are not too extensive. If the group can't agree on a design for the Extended One Generation Reproductive Study (EOGRTS), a revised version will nevertheless be developed including technical issues solved at the meeting. A report on the progress of this activity will be submitted to the WNT and the WNT will be requested to take a decision on the structure of the draft TG. In this case, another expert meeting might need to be envisioned.

5. The Chair added that participants will have to make a choice between principles (i.e. always test all possible end points within the reproductive cycle to exclude missing any effect) and pragmatism (i.e. save animals by excluding 2nd generation end points, when a 2nd generation is unlikely to contribute to Risk Assessment (RA) or Classification and Labelling (C&L)). The

objective would be to limit the parameters assessed in each cohort to essential components for making definitive determinations regarding particular endpoints (e.g. DNT) and avoid duplication with other TG. The group should also avoid reference to regulatory frameworks but design a scientifically sound and practical reproductive toxicity study.

Extended One Generation Reproductive Toxicity Study: Retrospective analysis

6. Aldert Piersma and Elizabeth Mendez presented the results of the retrospective analysis performed respectively in the NL and in the US.

7. The NL retrospective analysis, covering 176 studies was performed on industrial chemicals. The analysis indicated that the second generation neither had an impact on overall NOAEL nor on C&L. In some studies, findings in the F1 adults had an impact on NOAELs and on C&L. Thus the retrospective analysis supports the development of an EOGRTS. However, it was noted that the analysis may be biased by the availability of data and a comprehensive retrospective data analysis of 2-generation study results is advised to optimize the design of the EOGRTS and its use in the strategy for reproductive toxicity testing.

8. The US retrospective analysis was performed on 166 pesticides and 9 industrial chemicals. For reproductive effects, 2% of the studies showed a F_1 LOAEL > F_2 LOAEL and 1% had reproductive effects in F_2 generation only. For offspring effects, 5% of the studies showed a F_1 LOAEL > F_2 LOAEL and 4% had offspring effects in F_2 generation only. The conclusion was that F_2 generation has little value for establishing Reference Doses and for identifying effects unique in F2. The analysis also investigated in which cases the triggers proposed in the draft TG would have been used to lead to a second generation. The analysis showed that an F_2 generation would have been triggered for approximately 43% of the chemicals and that there was only one chemical (fenbutatin oxide) where the F_2 effect was the most sensitive of all generations and would not have been triggered.

9. Canada informed the participants that a retrospective analysis was also conducted in Canada, mostly on pesticides and that the results would be available shortly. As the analysis has been conducted on pesticides, there might be overlap with the database used by the US.

10. There was still some skepticism about the level of confidence that a study that would be terminated at the F1 generation would bring. Some participants were of the opinion that for classification, the F2 generation is often needed to confirm the finding in F1 and that it's not easy to classify only on the basis of F1 findings. When a 2 gen study is available, people involved in classification look at all data they have. It was indicated that in the NL data analysis, the critical effect had been used to conclude that the F1 would have sufficed for classification purposes.

11. It was also noted that the retrospective analysis performed could present a bias as we only have access to the studies performed on chemicals that have been marketed. Some chemicals might not be marketed because of reprotoxic properties, among them, some (how many?) might have effects only observed in F2. These chemicals would never enter data bases to which regulators have access. Ed Carney (BIAC) indicated that he didn't think that his company has chemicals that they didn't market because of severe effects observed only on the 2nd generation.

Two options for an Extended One Generation Reproductive Toxicity Study:

12. Ralph Cooper (US) presented the design of the EOGRTS to the participants. The protocol is based on the proposal of the ILSI/HESI ACSA protocol for testing food-use pesticides, as published in Cooper et al. in 2006. It was reminded to the group that in this protocol, the only triggers are those for production of an F2 generation. Cohorts 2 and 3 on developmental neurotoxicity (DNT) and developmental immunotoxicity (DIT) are not triggered and are part of the study design as proposed by the lead countries in version 5 of the draft TG (Meeting Document 1).

13. At the September 2008 meeting of the European Teratology Society (ETS), in Edinburgh, Chris Willoughby (UK) presented an alternative approach to the reproductive part of the EOGRTS, in a protocol including exposure from pregnancy of P females until mating of F1 and termination of the study at mid pregnancy (this protocol was referred to as the pre-post natal approach or PPN). The approach received a great support at the ETS meeting. The steering group considered it is worthwhile consideration and should be discussed at the meeting. Chris Willoughby was invited to present this approach at the expert group meeting. The objective of the PNN design is to address reproductive parameters as much as possible in the F1 generation. Parental generation (P generation) exposure is initiated during pregnancy, and F1 are mated by default and terminated midpregnancy.

14. Chris Willoughby also stressed the advantages of the PPN protocol, indicating that, efforts are not duplicated there (focus on F1 generation instead of P and F1) and it makes a more efficient use of animals, while continuing to ensure that all aspects of reproduction performance are tested. Whereas a complete reproductive cycle is tested in EOGRTS only if we go to the F2. Furthermore, mating outcome assessment in the second generation is an essential aspect of testing reproductive performance, which will be lost in the EOGRTS in case F2 is not triggered. It was also added that this concept can fit with the concept of modular design.

15. The two designs, PNN versus EOGRTS version 5 were discussed at length by the group in plenary session and in the reproductive breakout group. The main elements taken into consideration were the following:

Practicability

16. Some participants were concerned that internal triggers for a second generation in the EOGRTS may pose a significant problem of logistics (e.g., studies in the UK are planned 6 months in advance). With this protocol and triggering strategy, it is necessary to assume that a second generation will be required and to systematically plan personnel and animal facilities for the production of a second generation in case it is triggered. Whereas the PPN design would be more practical as it includes no triggers. With the PPN design, the mating of the F1 animals would be mandatory.

17. Furthermore, in an extended one-generation study, with synchronous allocation of F0 (original P) generation, all F1 offspring will be born within a 5 day period, making it very difficult to manage the additional complex evaluation required by DNT and DIT evaluations. This won't be the case with the PPN design.

18. People having experience with the protocol were invited to give their comments on practicality of the EOGRTS, including triggering of an F2 generation. The general conclusion was that it's not the easiest study to perform and might be logistically challenging but it's manageable.

Start of exposure

19. In the PPN design, dosing starts early in gestation (e.g. Gestation Day (GD) 4-6), whereas the EOGRTS requires a 2 week premating exposure. Some participants were concerned that with a starting point for dosing on GD 4 or 6, the adaptation to the chemical in the diet during pregnancy might affect the study outcome. This loss of weight due to diet palatability will be very difficult to interpret if exposure starts during gestation and pre-mating exposure may resolve this issue. It was noted that the PPN study design is regularly used in the pharmaceutical area, where gavage dosing was said to hardly ever give rise to such problems. As the new guideline is contemplated for pesticides and for industrial chemicals, dietary exposure will usually be the rule, and anticipated to give rise to the adaptation issues mentioned.

20. It was suggested that exposure could start at GD 0, in this case, all animals have adapted at GD6, where organogenesis starts. However, several participants felt that to avoid effects linked to stress/palatability and the favourability of steady state to be reached at the time embryogenesis starts, pre-mating exposure was considered advantageous.

21. In the EOGRTS, as exposure starts before mating, there is a potential loss of F1 offspring at the highest dose, which would preclude any F1 generation parameters studied as well as a subsequent F2 to be generated at that dose. However it was noted that if a prior dose range finding is done, dose selection will be well-informed, and this concern should be limited.

Animals savings and fertility assessment

22. In the PPN, the exposure starts during gestation, thus enabling not to count the P males in the number of animals used. The ethical question of the count of the F2 animals was debated: should the foetuses be counted as animals and from which day of pregnancy?

23. The PPN design might raise a problem of communication regarding animal saving. In different countries different legislation exists as to from what developmental stage a conceptus is regarded as an animal. Public perception also needs to be taken into account. Richard Vogel (Germany) and ICAPO were of the opinion that this protocol is not different from a 2 gen study. Given this discussion, it was felt that the default mating of the F1 in the PPN design would meet with significant regulatory and public resistance.

24. Some participants felt that scientific wishes should be separated from animal welfare considerations and that fertility is a major endpoint to evaluate, that can be done via mating of the unique F1 generation in the PPN design and not in the EOGRTS if F2 is not triggered. It was however noted that in the EOGRTS, fertility is assessed in P animals that are exposed via the diet during 2 weeks pre-mating.

25. It was suggested that F1 might be mated and terminated at observation of sperm or plug in female as a default in the EOGRTS. This would permit evaluation of mating behaviour of F1 without producing the F2 animals. Or that F1 could be mated to produce F2 for developmental aspects.

26. The general conclusion was that there was not a strong feeling from the group on these additions to the EOGRTS protocol. Neurotox experts indicated that if there is to be a positive result in mating behaviour of F1, where F1 is bred, it is likely that there would be other positive results in the neurotox cohort.

Male adult exposure

27. The group believed that the same effects would be picked up with the 2 designs, however, US has a concern with the use of the PPN design as from a regulatory point of view, there is a need in the US to differentiate between reprotox/fertility effects on young and on adult animals and thus a need to expose naïve adult animals (P generation in EOGRTS). The same applies for sperm analysis that needs to be done in adult naïve males, which will not be possible with the alternative protocol because all the evaluations will be done in the F1. It was however noted that this was a regulatory argument more than a scientific argument and that if a 28 day study is performed, it should be able to pick up effects on the adult male.

Validation

28. Although there is experience with the alternative protocol with pharmaceuticals, several participants were not sure what kind of validation would be needed with this protocol and there was some concern that this issue might delay the production of a Test Guideline.

29. EC/ECVAM indicated that the retrospective analyses performed for the EOGRTS are fine from the validation point of view (except from their point of view in the case of the DIT module). Parameters that trigger an F2 have been validated, e.g. vaginal opening or preputial separation. Whereas ECVAM didn't have time yet to review the PPN protocol from a validation perspective.

30. It was also noted that historical controls won't be available with the PPN protocol.

Consequences on cohorts 2 and 3

31. In the PPN design, cohorts 2 and 3 can be included as well. The experts in immunotox and neurotox were asked if exposure with the alternative design would make a difference regarding the results and data interpretation in DIT and DNT modules. The general feeling was that the results would be the same with either of the two designs. The PPN would have an advantage as it would enable to assess breeding behaviour outcome in F1, furthermore, it would be more manageable for planning neurotoxicity studies, because of staged mating.

Conclusion

32. Recognising that there are pros and cons for both protocols, the chair proposed an impromptu vote on the two protocols. There was an exact balance between countries/other bodies represented, although a numerical majority of experts supported the EOGRTS.

33. ⇨ The group considered the time frame for the development of the guideline and the necessary speed to introduce the guideline in order for it to be able to be implemented in the European REACH legislation, where it could impact highly on testing time, cost, and animal use. In view of this situation, and given that the EOGRTS proposal has already been widely discussed and finetuned in the past three years, the group agreed to continue discussing the EOGRTS draft, even if there was not a consensus in favour of the EOGRTS or the PNN protocol. EC/ECHA thought it was important to consider the outcome of the combined retrospective analysis (see para 55-64) in the future discussions.

Cohorts 2 and 3: should they be mandatory or not?

34. Before the start of the breakout group session, the Chair clarified that cohorts 2 and 3 are mandatory according to version 5 of the draft TG, which was unclear from the version 4 of the draft TG (Meeting Document 4) circulated to the expert group and the WNT in June 2008. Ralph Cooper confirmed this objective and also indicated that it is also the case in the ILSI/HESI ACSA

protocol. The Secretariat clarified that the last paragraph of appendix B, dealing with triggers for the immunotox cohort should have been deleted in Meeting Document 1.

35. EC/ECVAM, BIAC and ICAPO stressed that the change in the concept of triggering aspects between version 4 and version 5 of the draft TG constitutes a significant policy change and asked on which basis the changes were made. They questioned the transparency of the procedure and were surprised that this approach was changed without scientific consulting of the Expert group, especially as Industry currently is performing a series of feasibility studies based on the former draft versions

36. The Secretariat explained that version 5 (Meeting Document 1) is a proposal from the lead countries, which is brought to the experts group for discussion. This proposal was developed after receipt of comments from the expert group and the WNT on version 4 and after several conference calls of the lead countries. These conference calls also involved immunotox and neurotox experts from the lead countries. On the basis of the outcome of these conference calls and the comments received on version 4 of the draft (Meeting Document 4), the Secretariat revised the draft TG and prepared version 5 of the draft TG (Meeting Document 1). The comments received were also used to develop a list of technical issues for discussion during the meeting. The Secretariat indicated that for a better transparency, the summary record of the lead country conference calls will be posted on the protected website of the expert group meeting after the meeting.

37. The Chair suggested that a general sentence could be added in the TG, e.g.: "For hazard assessment of Agricultural chemicals, cohort 2 and 3 will be required. For other chemicals, cohort 2 and 3 are not necessarily mandatory but may be applied dependent on the regulatory environment. For all chemicals, an F2 generation will be triggered in exceptional cases only (to be specified)." This suggestion was heavily debated as detailed below, and no consensus was reached.

38. Some representatives from Germany, BIAC, ICAPO and EC/ECVAM emphasized the fact that there should be flexibility in the design to be adapted to different regulatory regimes, as indicated in the standard project submission form (SPSF) submitted to the OECD Secretariat in 2007.. They didn't agree that the cohorts should be mandatory in every case, for all types of chemicals and disagreed in particular with changes to paragraph 1 as well as the deletion of the last sentence of paragraph 31 from version 4 to version 5 of the draft TG. On the other hand, experts of some representing countries were in general in favour of having the cohorts 2 and 3 as mandatory cohorts and felt that the TG would be incomplete if they are not done. This subject was debated at length but no consensus could be reached. Some participants suggested that the issue should be discussed by the WNT. The following pros and cons were expressed:

39. Arguments in favour of a mandatory inclusion of cohorts 2 and 3:

- Cohorts 2 and 3 are part of the integrity of the TG, otherwise, we just have a classical (or even shortened – due to the decrease in the pre-mating exposure period of the P generation) one generation reproductive toxicity study, sometimes triggering an F2. The spirit of the study is to focus on different aspects of an exposure during the development.

- The study will be used for RA and C&L (see outcome of the breakout group sessions - para 43 to 54).

- Conducting cohorts 2 and 3 could avoid the conduction of further studies (e.g. TG 426).

- Neurotoxicity data are needed to address classification.

- In cases where developmental neurotoxicity and immunotoxicity data are not required by regulation it would be a pity to risk missing these effects by not addressing them, whereas this study offers this possibility.

- Even if other studies have been performed before a reproductive toxicity study starts, there are not many external triggers that could be useful to trigger a DNT or a DIT module.

- Some effects only appear if an organism has been exposed during development (e.g. immunotoxicity of TCDD).

- The conduction of the DIT and the DNT modules won't use more animals. Animals that are used in these cohorts are F1 animals that have been produced anyway by mating the P animals.

- Mutual Acceptance of Data (MAD) is easier to reach if the study is kept in its integrity.

40. Arguments in favour of flexibility in the conduction of cohorts 2 and 3:

- To meet the different global regulatory needs, the F1-extended one-generation study guideline needs to provide flexibility for different sector groups - therefore cohorts 2 and 3 should not be mandatory when neurotox and immunotox data are not required by the legislation.

- OECD TGs are designed to *accommodate* regulatory requirements, not *create* them.

- In order to address the common goal of reduced/refined animal testing the guideline should provide a more efficient testing approach by stressing the importance of starting with existing knowledge and using in-life observations to guide and tailor the testing.

- Performance of cohorts 2 and 3 would increase the potential for additional animal usage in further studies that may be triggered by type I errors (false positive results) in the main study, e.g. in the case of interference between endpoints concerning dose selection.

- Despite general agreement among experts that "a positive is a positive and a negative is a negative" for any given endpoint (e.g. DNT), one country present indicated that results of the EOGRTS may not be treated as definitive, but rather as a trigger for TG 426.

- Although it have shown to be manageable, inclusion of cohorts 2 and 3 as mandatory cohorts would result in a complex design for routine evaluation of industrial chemicals.

- The performance of cohorts 2 and 3 in all cases will increase the cost per substance significantly.

- Some of the functional assays in the DIT module have not been sufficiently validated according to the specifications of OECD GD 34.

- The Standard Project Submission Form (SPSF) initially submitted by the lead countries reflects flexibility in the application of cohorts 2 and 3.

41. ⇨ It was decided to bring this point of discussion to the WNT. To prepare for the WNT meeting, the Secretariat will send a letter to the WNT after the meeting and ask for its views.

Report from the breakout group session

42. Half a day during the meeting was dedicated to breakout group sessions. The group was divided into 3 breakout groups to discuss (1) the reproductive module, (2) the DNT module and (3) the DIT module. Experts specialized in neurotox and immunotox were invited to join their respective group, other experts could attend the session of their choice.

DNT module

43. The majority of members of the group on the DNT module was in favour that a limited DNT module should be included as a default in the study. Their concern was that an EOGRTS might be avoided because of risk of neurotoxic effects, which might lead to classification, labelling and more testing. However, it was noted that the decision on the performance of the study is not entirely a business decision but also depends on regulatory requirements.

44. The group felt it was important to balance testing requirements and to preclude overloading the EOGRTS. This is an issue for both acceptance/use and practicality. The group discussed the possibility to add more than the current proposal, in particular, sexually dimorphic behaviours and learning and memory testing. In plenary session, the whole group agreed that learning and memory tests would not be included as a default in the EOGRTS, as they are time consuming but opening would be left in the TG; there would be a possibility to do it, in particular if there is an indication from previous research.

45. The group discussed how to interpret the data. The neurotox group felt that a negative results in the DNT module should not necessarily lead to further testing and asked the whole group what should be the follow-up in case of positive results in the DNT module. Should this be taken as a starting point for RA and C&L or as a starting point for more work? The group suggested that a Guidance Document could be developed in support of this TG, in particular to address interpretation of data and the need for further testing in case of positive or negative results.

46. The Chair suggested that if there is an effect, this may be considered as adequate information for RA and C&L and if the test is negative, the content of the DNT module provides enough elements to have confidence in the results and there's no need to go to a complete DNT study. The majority of experts agreed with this approach. Neurotox experts agreed that there is little chance that a chemical that gives negative results in the DNT module as proposed would give positive results in a complete DNT study.

47. ⇨ The conclusion of the group was that, although this limited DNT testing cannot be interpreted as a replacement for a DNT study if DNT data is required by regulation, the DNT data may be sufficient for regulators, assuming that a positive result is taken as a positive and a negative result is taken as a negative. The DNT results could then be used by regulator in decision making regarding C&L and RA. The need for a guidance document on data interpretation was expressed by the participants. The subgroup proposed that the DNT module should consist of motor activity and brain morphometrics, and this was generally accepted.

DIT module

48. The immunotox group concluded that scientifically, the best way to assess immunotoxicity is a functional test. Therefore, the proposal of the immunotox group was that the assessment of the primary IGM Antibody response to a T cell dependant antigen (TDAR) should be performed in

cohort 3. As it is the most validated immune functional endpoint, it was the only one that was proposed to be included in the DIT module of the draft TG.

49. The group also recommended to include evaluation of some parameters in cohort 1 for comparison with other parameters measured in the same animals and also because animals in cohort 1 are not immunized, whereas animals in cohort 3 are immunized. In addition to parameters already included such as haematology, clinical chemistry, organ weight and histopathology, the group recommended inclusion of phenotypic analysis of splenic cell types (incl. B-cells, total T-cells, T-cell subsets and NK cells) and assessment of bone marrow cellularity.

50. ICAPO noted that some of the functional immunotox assays listed paragraph 56 in the draft TG may not be considered sufficiently validated according to OECD GD34, and urged immunotox experts to revise the 2-page overview that was prepared in advance of the expert meeting (referenced as INF 9 in the list of documents available for the meeting), to provide a more robust description of studies undertaken to date which demonstrate the relevance and reliability of the methodologies being proposed.

51. Bob Luebke (US) will revise the document and provide more details on TDAR validation studies. Regarding phenotypic analysis, its use in immunotoxicology has been validated in mice and rats. The immunotoxicology experts cited numerous studies that have clearly demonstrated heightened sensitivity to a wide variety of chemicals during immune system development and maturation. Examples will be provided in the TDAR validation document. Although formal guidelines for conducting DIT studies have not been developed to date, resulting in a variety of exposure and functional testing designs, studies conducted in multiple laboratories have proven that a variety of experimental designs are successful in detecting developmental immunotoxicity. The greater sensitivity of the developing immune system to various chemical classes, the widely accepted superiority of functional testing to establish effects on the immune system compared to observational endpoints, and the validity of the current TDARs to establish immunotoxicity indicate that in fact there is quite a solid data base for functional testing in this context available.

52. ⇨ The group suggested that the same conclusion of those on the DNT apply to the DIT module in term of interpretation of results: take a positive for a positive and a negative for a negative and that the inclusion of this module in the TG should then not lead to further immunotoxicity studies and should be used for C&L and RA. A treatment-related increase in the TDAR should be considered an adverse effect. Immunotox experts agreed that a minor effect might not be relevant, but if there several parameters show similar changes, then the result should be regarded as positive. Weight of evidence approach should be used, in the context of RA. Here too, the group recommended the development of a Guidance Document giving more details on data interpretation than in the TG.

53. For regulatory context in which a DIT module would not be a presumed default, the group indicated that it should be triggered on the basis of positive immunotoxicologic findings in previous general toxicity studies. However, it was questioned whether e.g., OECD TG 407 would be sufficiently sensitive to provide relevant triggers.

54. The group was confident that an 'N'=10/sex is sufficient. Hence there will be no need for the second set of 10 males and females for cohort 3 which means a further reduction in animal numbers. The group also agreed to have more flexibility on the day when the functional test should start, e.g. day 50-70.

Reproductive module: triggers for an F2

55. Two issues were discussed by the breakout group on reproductive toxicity (i) the value of the F2 generation and (ii) when an F2 should be triggered.

56. The retrospective analyses have shown that F2 are not often necessary for RA and C&L. However, the following questions were asked: How to measure the risk that would be taken in terminating the study at F1? Is it justified to take additional risk? It was noted that whatever protocol used, there will be residual risk. Our RA might already include risk of false negatives. If this draft TG is adopted, we will exchange one suboptimal option for another one. It was however noted by EC/ECHA if the production of a 2nd generation should be skipped, there is a need to have really strong arguments to know what will be missed. Others pointed out that the F1 end points in the EOGRTS are more numerous and more sensitive than the F1 end points in the two-generation study, whereas the F2 end points in both study designs are essentially identical. Therefore, the results of the retrospective analyses of past two-generation studies probably overestimate the risk of missing a reproductive effect.

57. Some participants noted that it would be useful to look at the chemicals in the 2 retrospective analyses that gave an effect in F2 but not in F1. There is uncertainty regarding these cases. It might be that specific findings in an F2 are a question of chance, or there might be a specific mechanism, or F2 might be just a little more sensitive than F1. Answers to these questions might help in choosing an optimal protocol.

58. ⇨ The group agreed that there was a need to merge the databases available (US, NL and Canada), to go back to the data analysis, compare F1 and F2 and try to respond to the questions raised in the paragraphs above. It was also requested that any other data available should be brought to the knowledge of the group. The Secretariat will send a letter to try to gather other information available. BIAC (BASF) proposed to provide their comparison between EOGRTS and TG 416 for vinclozolin as the in life studies have already been done. Other companies will shortly perform other studies (Syngenta, Bayer Cropscience).

59. If the triggers proposed in the draft TG are used, the US retrospective analysis showed that the triggers would be used - and thus the study would go to the 2nd generation - in 43% of the cases. As the results obtained by the NL retrospective analysis were close to the US one, it could be expected that for industrial chemicals, the percentage of use of the triggers would be in the same range.

60. Concern was expressed that the triggering of an F2 within the study was not practical, both in view of necessary reservation of animal rooms and personnel for an as yet uncertain part of the study, and in view of the question whether data that should provide the triggers could be analysed in time to be able mate the F1 before they would be of too advanced age. Triggers from earlier performed studies would not have this disadvantage.

61. ⇨ The group agreed that the triggers, as defined in the current draft protocol, may be too many, too sensitive and too conservative and that there would be a need to refine them in the draft TG, so that they can better focus on the identification of relevant cases.

62. ⇨ Timeline: the group agreed on a 2 step process: in the short term, start with the TG we have and submit it to the WNT for adoption at the next WNT meeting, in April 2009 with the caveat that triggers need to be revisited. Then, refine the TG on the subject of F2 triggering within one

year. In parallel a new analysis of data would be performed. It could be foreseen that a TG with refined triggers could be submitted to the WNT for final agreement in 2010.

63. ⇨ To conduct this further data analysis, the group agreed that the available data bases (US, NL, Canada) should be combined and that the three lead countries and Canada would work together on the refinement of the retrospective analysis. The EC/ECVAM and EC/ECHA indicated that they were prepared to consider supporting such a data analysis. It was also agreed that there should be another expert group meeting to discuss the results of the new analysis when available (fall 2009).

64. ⇨ It was agreed that for a thorough examination of the data and triggers, all relevant questions to consider in the data analysis should be collected. The participants were invited to send any relevant questions to the Secretariat.

Technical issues

65. Toxicokinetic studies: This issue was discussed in plenary session and by the breakout group on reproductive toxicity.

66. ⇨ There was a consensus that TK studies should be performed as part of a dose range finding study, prior to the main study. This would help for setting dose levels and also for having information on placental and milk transfer prior to the study. The group also agreed that there would be an option of taking samples during the study at various time points, to check for internal dose. The range finding study should be directly linked to the study and use 4 litters as there are 4 time points. It was also noted that it is important that the steady state is reached before sampling.

67. Later in the course of the meeting the group reworded the paragraphs related to TK studies to reflect this consensus. During the discussion, the group agreed to have only 3 time samplings (instead of 4) – it was agreed that to assess exposure via milk, take a sample once (on PND 10) instead of twice (on PND 4 and PND 20/21) was enough.

68. Number of pups per cohort: Canada had a concern with the use of 1 animal/sex/litter in cohort 1. From a regulatory point of view, Canada will not accept the EOGRTS for pesticide registration if only 1 animal/sex/litter in cohort 1 is used. 5 animals/sex/litter are kept; from these 5, 1 is used for DNT and 1 is used for DIT, thus the Canadian proposal is that 3 animals/sex/litter are kept for cohort 1. They argued that by doing this, the statistical power and robustness of the data would be increased. Some participants objected that the statistical power of the study would not be increased a lot as the litter is the statistical unit, and furthermore, it would require a lot of extra work for a low gain. Other participants felt it was worthwhile from a scientific point of view, in particular as it would provide more data and help in detecting rare malformations. However, it was stated that for the latter, necropsy at weaning would suffice in most cases.

69. ⇨ As there was no consensus on this issue, it was agreed that scientific arguments on this would be exchanged after the meeting, via the Secretariat and the issue should be solved within one month time. A conference call will be organized on this issue once comments have been received and Canada had internal discussion. If there is no consensus between experts, the questions will be forwarded to the WNT.

70. Culling issue: The issue of culling was debated at the meeting. Beate Ulbrich (Germany) had provided an issue paper on culling before the meeting (INF 2). US, Canada and UK were in favour of culling. The other European countries were not in favour of culling. For several parameters there are no data, to date, indicating that culling or not culling has an influence on results. Some participants were concerned that for laboratories that are used to cull, to stop culling might induce variations from historical controls. It was noted that standardization of litter size to 10 might not result in terminating many pups, as 10 might be very close to natural litter sizes. It was agreed to keep this item flexible in the guideline. The original paragraph 29 was kept. This paragraph provides flexibility regarding this issue.

71. Pup dosing: The question of dosing the pups before weaning in the cases where there is not milk transfer was discussed. Some experts felt here it's not necessary as it stresses the pups. Furthermore some felt that the benefit of attempting gavage is less than the challenge of carrying it out logistically. The question of human relevance was also raised. Other experts were in favour of dosing the pups when there's no compound transfer to milk, otherwise, pups may not be exposed post-natally. This is of concern as the development in rats is done earlier than in humans and the lactation period in rats corresponds to the last trimester of human pregnancy. In conclusion, paragraph 22 was kept as it is as the group felt it gives enough flexibility.

72. Inclusion/exclusion of runts: The group agreed that the runts should not be excluded completely of the study, as they represent an effect and they should be submitted to gross necropsy. However they should not be used in the reproductive analysis as they are not representative of their litter. It was also pointed out that keeping obvious runts that are unlikely to survive more than a few days would ensure inadequate offspring for offspring testing. The number of runts in a litter or the number of litter including runts should be counted and documented.

73. Thyroid hormone measurement: One of the commenters on version 4 of the draft TG questioned the usefulness of measuring thyroid hormones as it is technically challenging and might be a confounding factor. The group argued that these parameters have been used clinically in hospital for 40 years. Furthermore, a report on validation has been developed and should be shortly available. The group agreed to keepT4 and TSH measures in the protocol.

74. The US indicated that a survey on mammary gland is being performed and should be shortly available.

75. Organ Weights/Tissue Preservation: There were discrepancies in this section and confusion about the tissues/organs which should be weighed and preserved, vs. those suitable only for preservation. It was agreed that this section needs to be revised. Ed Carney offered to provide a revised draft of this section after the meeting.

Conclusion

76. To wrap up the meeting main conclusions, the chair prepared 2 slides on the meeting outcomes and timelines (provided below). The draft Test Guideline (version 5 of the EOGRTS) was revised during the meeting on the basis of the discussions and agreements reached by the group. It is attached as an annex to this summary record. BIAC indicated that in order to meet the OECD tenet of transparency the draft needs to be circulated to a broader scientific community.

77. Once the summary record is agreed by the participants, it will be sent to the WNT with a letter pointing out the issues where the group couldn't reach consensus. At the same time, the draft TG revised at the meeting will be circulated to the WNT for comments.

Meeting outcomes (slides as shown at the meeting)

1. Starting point for further development will be the ext-1-gen study design (updated draft 5), derived from the original ACSA proposal.

2. Agreement was reached on the contents of cohorts 2 and 3. They were defined as the minimal essential package needed for a risk assessment decision on developmental neurotoxicity and developmental immunotoxicity, respectively. This package was considered as a practically feasible package in view of carrying out the study.

3. The triggers for the F2 generation will have to be revisited as it was felt that the current guidance would possibly lead to triggering an F2 in too many cases.

4. To guide reformulation of the triggers for the F2, the current data analyses by NL, US and Canada should be merged and appropriate questions should be asked to the entire dataset. (Questions will be collected from the meeting participants through OECD secretariat). In addition, OECD will be asked to put out a request for any existing data that could be relevant in view of the relative sensitivity of the F1 versus the F2 generation. ECVAM and ECHA will consider supporting this activity.

5. The Canadian representatives will readdress their need for assessing 3 versus 1 pup per sex per litter for cohort one at necropsy, and report back in 30 days. (comments from the meeting participants to Canada ASAP).

6. A text proposal in view of Mutual Acceptance of Data was discussed at length and gave rise to major differences in opinion. Therefore, this item was referred to the WNT. The specific arguments brought forward will be listed in the minutes of the meeting.

Time line

20 Nov 2008	Canada to report back on 1 vs 3 pups per sex per litter issue
End Nov 2008	Teleconference on consequences and finalizing text of draft guideline and submission to WNT with caveat that F2 trigger text will be readdressed within one year (for WNT 2010)
March 2009	WNT to consider draft guideline with current F2 trigger text
Fall 2009	Finalized integrated data analyses on F1 vs F2 sensitivity issue in retrospective 2-gen study data analysis
Nov 2009	Expert group meeting on data analysis outcome and generation of novel text for F2 trigger guidance
March 2010	WNT to discuss draft plus new F2 trigger text

Annex

Draft TG on an EOGRTS, as revised during the meeting

Track changes of version 5 have been accepted, so that this version only shows changes accepted at the meeting

Merged draft version ~~5~~6, 2008-10-~~03~~17

~~**Text highlighted in yellow raises particular discussion points**~~

Preliminary statement:

This project is developed by a team of 3 lead countries, the USA, Germany and the Netherlands. The present draft TG reflects the consensus of the 3 lead countries for many issues but NOT for all of them. Some of the comments that have been inserted in the draft focus on these issues, still debated within the lead countries.

DRAFT Extended One-Generation Reproductive Toxicity TEST GUIDELINE

OECD/OCDE XXX

OECD GUIDELINE FOR THE TESTING OF CHEMICALS

INTRODUCTION

1. This Test Guideline has been designed to address the common goal of reduced/refined animal testing for toxicity to reproduction and offspring development, while preserving the integrity of the science. It is based on the International Life Science Institute (ILSI)-Health and Environmental Sciences Institute (HESI), Agricultural Chemical Safety Assessment (ACSA) Technical Committee proposal for a life stage F_1 extended study published in Cooper et al., 2006 [1]. Several improvements and clarifications have been made to the study design initially proposed by Cooper *et al.,* to provide flexibility and a more efficient testing approach by stressing the importance of starting with existing knowledge and using in-life observations to guide and tailor the testing. This guideline provides a detailed description of the operational conduct of an Extended F1 One-Generation Reproductive Toxicity Guideline. This TG also lists a set of endpoints that would be used to determine the need to produce an F2 generation. The TG

describes three cohorts of F_1 animals, Cohort 1 assesses reproductive/developmental endpoints Cohort 2 evaluates potential impact of chemical exposure on the developing nervous system and Cohort 3 used to evaluate possible effects on the developing immune system. Cohort 1 may be extended to include an F_2 generation, depending on available data including observations during the in-life part of the study.

INITIAL CONSIDERATIONS AND OBJECTIVES

2. In a battery for efficient toxicity testing the Extended One-Generation Reproduction Toxicity Study is used to detect effects on reproductive endpoints that are not covered in repeat-dose toxicity studies of 28 or 90 day duration, and effects that may occur as a result of pre- and postnatal exposure. For reproductive endpoints it is envisaged that, as a first step, use is made of the repeat-dose studies (including screening reproductive toxicity studies, e.g. TG 422, or short term endocrine disrupter screening assays, e.g. pubertal assay) to detect effects on reproductive organs for males and females. This would include spermatogenesis (testicular histopathology) for males and oestrus cycles, follicle counts/oocyte maturation and ovarian integrity (histopathology) for females. The Extended One-Generation Reproduction Toxicity Study then serves as a test for reproductive endpoints that require the interaction of males with females, females with conceptus, and females with offspring [1].

3. The TG is designed to provide an evaluation of the pre- and postnatal effects of chemicals on development as well as a more thorough evaluation of systemic toxicity and ADME in pregnant and lactating females and young and adult offspring. Detailed examination of key developmental endpoints, such as offspring viability, neonatal health, developmental status at birth, and physical and functional development until adulthood is expected to identify specific target organs in the offspring, including the reproductive, endocrine, neural, and immune systems. In addition, the study will provide and/or confirm information about the effects of a test substance on the integrity and performance of the adult male and female reproductive systems, with respect to gonadal function, the oestrus cycle, epididymal sperm maturation, mating behaviour, conception, pregnancy, parturition, and lactation [2]. The data derived from this test should allow the determination of No-Observed Adverse Effect Levels (NOAEL), Lowest Observed Adverse Effect Levels (LOAEL) and/or benchmark doses [3] for the various endpoints and serve as a guide for subsequent testing.

---

[1] **NOTE: Endpoints covered**

This includes mating performance and success (e.g.sperm maturation, behavioral parameters, secondary sex organ function), pregnancy (e.g. fertilisation, pre-implantation embryo transport to the uterus, priming of the uterine epithelium for implantation, placentation, hormonal support of pregnancy), parturition, maternal behaviour and lactation, pre- and post-implantation embryo differentiation and development, foetal development, adaptation to extrauterine life, and postnatal development and function.

[2] **NOTE: Confirmation and characterisation of previous findings**

This study may be used to characterise effects detected in previous repeat-dose studies. However, care should be taken to avoid an interference of these confirmatory investigations with the objective of obtaining sufficient litters at all dose levels. Depending on the effects observed in the repeat-dose study, specifically designed Tier 2 studies may often be more suitable for characterisation.

[3] **NOTE: Benchmark dose**

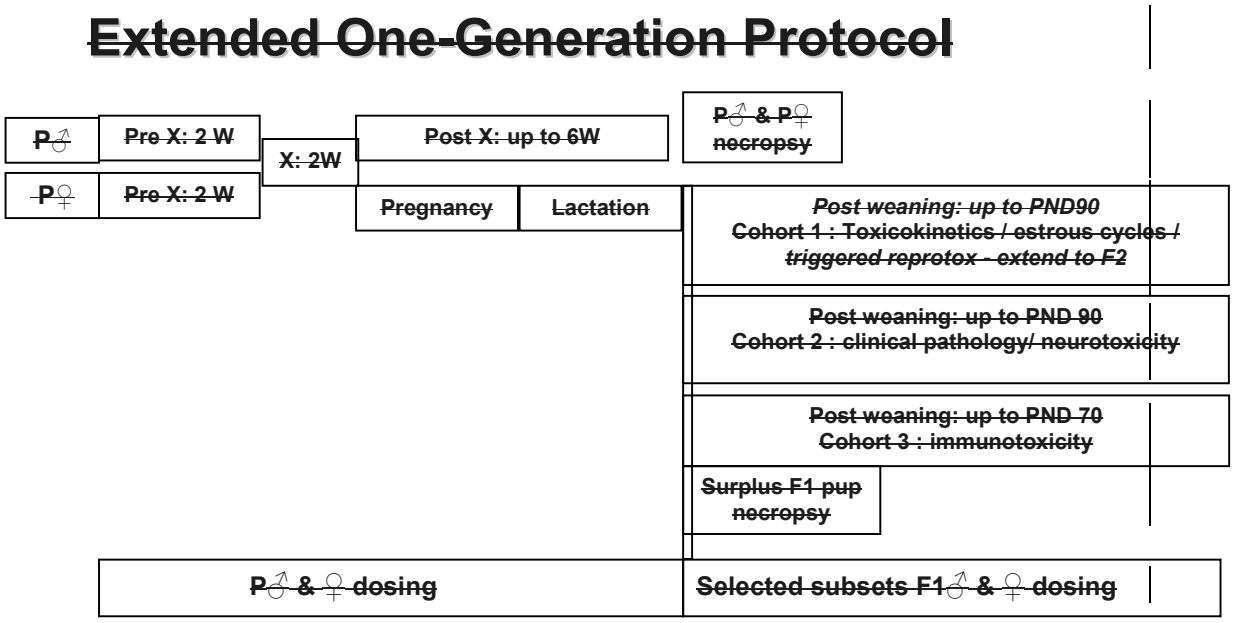
When the ~~available~~nature of the data do not allow clear determination of a NOAEL, the ~~preferred~~alternative approach to derive a point of departure (PoD) is benchmark dose (BMD) modeling. It should be noted that for BMD modeling, a benchmark response should be specified for each endpoint. Thus, BMD modeling may not always be feasible due to either the ~~quality~~nature of the data or the challenges in specifying a benchmark response for every endpoint. For full support of the BMD approach, studies may have to be designed specifically.

PRINCIPLE OF THE TEST

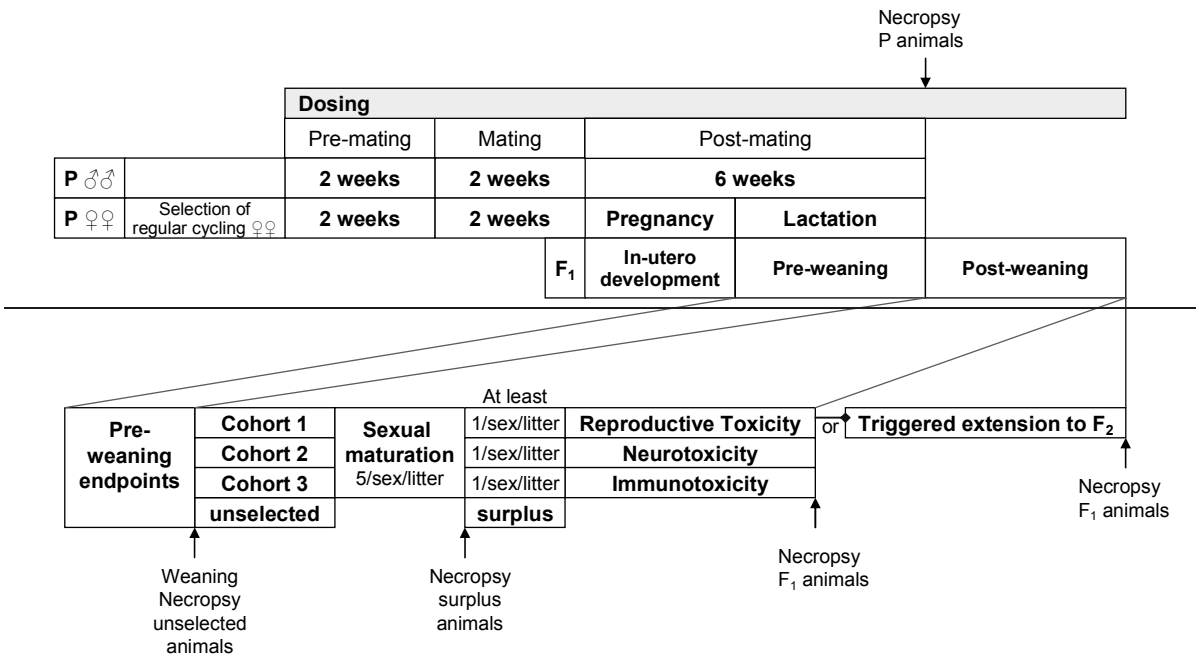
4. A schematic drawing of the protocol is presented in Figure 1. The test substance is administered continuously in graduated doses to several groups of sexually mature males and females. This parental (P) generation is dosed for a defined pre-mating period (selected based on the available information for the test substance) and a two-week mating period. P males are further treated at least until weaning of the F_1. They may be treated for longer if this is needed to clarify effects on reproduction. Treatment of the P females is continued during pregnancy and lactation until termination after the weaning of their litters. The F_1 offspring receive further treatment with the test substance from weaning to adulthood until postnatal day (PND) 70 or 90, depending on cohort assignment. If a second generation is triggered, The F1 offspring will be maintained on treatment until weaning of the F_2, or until termination of the study if the study is terminated before weaning of F_2.

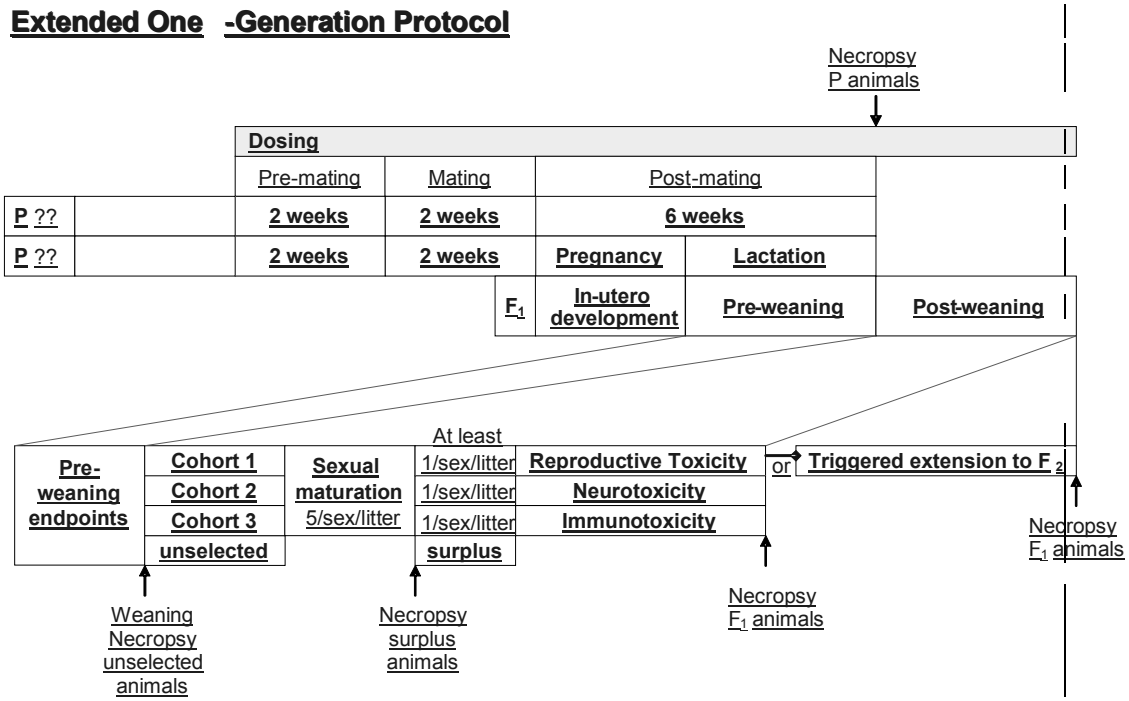
5. Clinical observation and pathology examinations are performed on all animals for signs of toxicity with special emphasis on the integrity and performance of the male and female reproductive systems and the health, growth, development and function of the offspring. At weaning, selected offspring are assigned to specific subgroups (Cohort 1-3, see paragraphs 30 and 31~~32~~) for further investigations, including (but not restricted to) sexual maturation and reproductive organ integrity, neurological and behavioural endpoints and immune functions.

Figure 1: Scheme of the Extended One-Generation Reproduction Toxicity Study



Extended One-Generation Protocol





DESCRIPTION OF THE METHOD / PREPARATIONS FOR THE TEST

ANIMALS

Selection of animal species and strain

6. The choice of species for the reproductive toxicity test must be carefully considered in the light of all available information. However, because of the extent of background data and the comparability to general toxicity tests, the rat is normally the preferred species. If other species are used, justification should be given and appropriate modifications to the protocol will be necessary. Strains with low fecundity or a well-known high incidence of spontaneous developmental defects should not be used.

Age, body weight and inclusion criteria

7. Healthy parental animals, which have not been subjected to previous experimental procedures, should be used. Both males and females should be studied and the females should be nulliparous and non-pregnant. The P animals should be sexually mature, as nearly as possible of uniform weight and age (approximately 80 days) at the initiation of dosing, and representative of the species and strain under study. It is recommended that delivery to the test facility occurs at 10 weeks of age for P males and at 8 weeks of age for P females. Animals should be acclimated for ~~10~~ at least 5 days after arrival~~. To prevent the inclusion of non-cycling females into the study, vaginal smears are obtained for a period of 14 days prior to dosing. Only females displaying regular 4-5 day estrous cycles should be selected for testing.~~.. The animals are randomly assigned

to the control and treatment groups, in a manner which results in comparable mean body weight values among the groups~~. At the commencement of the study, the weight variation of animals used should be minimal and not exceed 20 % of the mean weight of each sex~~.

Housing and feeding conditions

8. The temperature in the experimental animal room should be ~~21~~22 °C (+/- ~~2~~3°). Although the relative humidity should be at least 30 % and preferably not exceed 70 % other than during room cleaning, the aim should be 50-60 %. Artificial lighting should be set at 12 hours light, 12 hours dark. Conventional laboratory diets may be used with an unlimited supply of drinking water. Careful attention should be given to diet phytooestrogen content~~, and their concentration(s) should be reported.~~. Standardized, open-formula diet in which estrogenic substances have been reduced to levels that do not alter results of studies that are influenced by exogenous estrogens is recommended (48).The choice of diet may be influenced by the need to ensure a suitable admixture of a test substance when administered by this method. Content, homogeneity and stability of the test substance in the diets have to be verified. The feed and drinking water should be regularly analysed for contaminants. Samples of the diet should be retained until finalisation of the report, in case the results necessitate a further analysis of diet ingredients.

9. Animals should be caged in small groups of the same sex and treatment group. They may be housed individually to avoid possible injuries (*e.g.* males after the mating period). Mating procedures should be carried out in cages suitable for the purpose. After evidence of copulation, females that are presumed to be pregnant are housed separately in delivery or maternity cages where they are provided with appropriate and defined nesting materials when parturition is near (gestation day 16-18). Litters are housed with their mothers until weaning. Each cohort of selected F_1 animals is housed in small groups of the same sex and treatment group from weaning to termination.

Number and identification of animals

10. Normally, each test and control group should contain a sufficient number of mating pairs to yield at least 20 pregnant females per dose group. The objective is to produce enough pregnancies to assure a meaningful evaluation of the potential of the substance to affect fertility, pregnancy and maternal behaviour of the P generation and growth and development of the F_1 offspring from conception to maturity. Failure to achieve the desired number of pregnant animals does not necessarily invalidate the study and should be evaluated on a case-by-case basis, considering a possible causal relationship to the test substance.

11. ~~Toxicokinetic measurements should be made in at least 4 litters/dose/age group (1 pup per sex/litter) [4] either in a range-finding study prior to the main study or in satellite animals as defined elsewhere in this document.~~

12. Each P animal is assigned a unique identification number before dosing starts. P females monitored for oestrus cycles after the acclimation period should receive their identification before this procedure is initiated. All F_1 offspring are uniquely identified when neonates are first examined on postnatal day (PND) 0 or 1. Records indicating the litter of origin should be maintained for all F_1 animals, and F2 animals where applicable, throughout the study.

TEST SUBSTANCE

Available information on the test substance

13. The review of existing information is important for decisions on the route of administration, the choice of the vehicle, the selection of animal species, the selection of dosages and eventually also for modifications of the dosing schedule. Therefore, all available information on the test substance, i.e. physico-chemical, toxicokinetic and toxicodynamic properties, structure-activity relationships (SARs), results of previous toxicity studies (*e.g.* acute toxicity, toxicity after repeated application), and relevant information on structural analogues should be taken into consideration in planning the Extended One-Generation Reproduction Toxicity Study. Preliminary information on ADME and bioaccumulation may be derived from chemical structure, physico-chemical data and extent of plasma protein binding while results from toxicity studies give additional information, *e.g.* on NOAEL, metabolism or induction of metabolism.

Route of administration

14. Selection of the route should take into consideration the route(s) most relevant for human exposure. Although the protocol is designed for administration of the test substance through the diet, it can be modified for administration by other routes (drinking water, gavage, inhalation, dermal), depending on the characteristics of the compound and the information required.

Choice of the vehicle

15. Where necessary, the test substance is dissolved or suspended in a suitable vehicle. It is recommended, that, where possible, the use of an aqueous solution/suspension is considered first, followed by consideration of a solution/suspension in oil (*e.g.* corn oil). For vehicles other than water, the toxic characteristics of the vehicle must be known. Use of vehicles with potential intrinsic toxicity should be avoided (*e.g.*, acetone, DMSO). The stability of the test substance in the vehicle should be determined. Considerations should be given to the following characteristics if a vehicle or other additive is used to facilitate dosing: effects on the absorption, distribution, metabolism, or retention of the test substance; effects on the chemical properties of the test substance which may alter its toxic characteristics; and effects on the food or water consumption or the nutritional status of the animals.

Selection of Dosage

16. Normally, the study should include at least three dose levels and a concurrent control. All available data should be utilized to aid dose selection, including the results of systemic toxicity studies[5], and if available, metabolism and kinetics (in this case, particular emphasis should be given to avoid high dose levels which saturate kinetic processes). The dose levels should be spaced to produce a graded toxic effects. Unless limited by the physical/chemical nature or biological properties of the test substance, the highest dose should be chosen with the aim to

[5] **NOTE: Dose selection**

When selecting appropriate dose levels, the investigator should consider all available information, including the dosing information from previous studies as well as the toxicokinetics (TK) in pregnant or non-pregnant animals, the extent of transfer to the milk of the lactating female, and information on induction of metabolism or bioaccumulation of the test substance. These data will assist in demonstrating the adequacy of the dosing regimen. Compounds with short half-lives and milk concentrations less than or approximating the mother's blood concentrations will likely produce low internal exposures in the pups during lactation. Compounds that are highly excreted in milk (*e.g.*, milk: blood partition coefficient greater than approximately 2 or 3) and cleared with a half-life of approximately 24 hrs or longer may produce substantial internal exposures in offspring. Because of the high food consumption of lactating females and/or because of physiological differences in very young animals this may result in a higher effective dosage than in adults exposed to the same concentration. Food consumption in weanling pups (on a mg/kg/day basis) is approximately double that of adult animals, scaling allometrically as body weight raised to the ¾ power. TK data collected as part of a range-finding study would directly address these issues. TK data can also be useful for informing adjusted dietary doses. For example, if internal exposures in offspring are anticipated to exceed those of the adult animal based upon TK analyses, the investigator may consider reducing the top concentration of test substance in diet or water during the lactation period and during the early post weaning life-stages of the F1 generation so as to avoid excessive toxicity. If excessive maternal toxicity were anticipated during the lactation period, reducing the concentration of test substance in the diet or water during this period could be considered. It should be noted that any toxicokinetic data needs to be very robust before adjusting (decreasing) doses during any time of reproduction.

induce some systemic toxicity but not death or severe suffering in P males and non-pregnant P females. ~~However, some substances may cause unforeseen toxicity and death specifically in pregnant females through a mechanism related to the pregnant state. In such a case additional studies under an adapted protocol may be necessary to elucidate postnatal effects on offspring.~~ A descending sequence of dose levels should be selected in order to demonstrate any dosage-related effect and to establish no-observed-adverse-effects levels (NOAEL) or doses near the limit of detection that would allow to derive a benchmark dose for the most sensitive endpoint(s). To avoid large dose spacing between NOAELs and LOAELs, two- or four-fold intervals are frequently optimal. The addition of a fourth test group is often preferable to using a very large interval (*e.g.*, more than a factor of 10) between dosages.

17. Except for treatment with the test substance, animals in the control group are handled in an identical manner to the test group subjects. This group should be untreated or sham-treated or a vehicle-control group if a vehicle is used in administering the test substance. If a vehicle is used, the control group should receive the vehicle in the highest volume used.

Limit test

18. For substances with low toxicity (i.e. if a dose of at least 1000 mg/kg body weight/day produced no observable toxic effects in any systems in repeat-dose studies) or if toxicity would not be expected based upon data from structurally and/or metabolically related compounds, a full study using several dose levels may not be necessary and the Extended One-Generation Reproduction Toxicity Study could be conducted using a control group and a single high dose level of at least 1000 mg/kg body weight/day. However, should evidence for reproductive or developmental toxicity be found at this limit dose further studies at lower dose levels will be required to identify a NOAEL. These limit test considerations apply except when human exposure indicates the need for a higher dose level.

PROCEDURES

Toxicokinetics and exposure of offspring

19. To aid the planning of the study design, selection of dose levels and ~~the~~ interpretation of results, it is highly recommended that a minimal toxicokinetic ~~measurements could~~data set be ~~made~~available *a priori*, either from a dose range-finding study in ~~at least 4 litters/dose/age group (~~pregnant/lactating animals or equivalent data. The purpose of these data is: 1~~ pup per sex/litter)~~ to verify exposure of developing fetuses and ~~the corresponding P females)~~ [6]. ADME ~~pups~~ to the test compound (or relevant metabolites), 2) to provide a rough estimate of internal dosimetry ~~data could~~in order to assist in data interpretation and risk assessment, 3) to investigate potential non-linear, dose-dependent changes in kinetics. Additional toxicokinetic data, such as metabolite profiles, concentration-time courses, etc. should also be considered if they are available. Supplemental data may also be collected ~~as part of a range-finding study. The analyte to measure (parent substance and/or key metabolite(s) when possible) and~~ during the main study, provided that it does not interfere with the ~~sampling time~~ collection and interpretation of the main study end points~~ should be determined from the results of previous pharmacokinetic/toxicokinetic studies and/or repeat-dose toxicity studies. If a range-finding study using~~.

[6] ~~**NOTE: Toxicokinetic data**~~

~~Depending on the limits of detection and/or quantification for the test substance or relevant metabolite(s) pooled samples of individual litters may be used.~~

20. A recommended minimal data set which should be available prior to the ~~relevant route of exposure and assessing milk, amniotic fluid, and maternal/fetal/pup blood is not available, or~~ conduct of the ~~internal dose cannot be estimated from the available data, a toxicokinetic study could be conducted in a satellite population of animals.~~

~~20. When toxicokinetics studies are performed, recommended data collection time points include~~main study is as follows:

- Late pregnancy (*e.g.*, GD20) for maternal blood and foetal blood
- Midlactation (PND 10) for maternal blood, pup blood and/or milk
- ~~Early lactation (*e.g.*, PND 4) for maternal and neonatal blood samples[7], and maternal milk~~
- ~~Late lactation (*e.g.* PND 20/21) for maternal and juvenile/weaning blood samples~~
- Early post-weaning (*e.g.*, PND28) for weanling blood samples

21. ~~Collection~~The number and timing of ~~data~~sample collection on a given sampling day will be dependent upon route of exposure and prior knowledge of toxicokinetic properties in non-pregnant animals. For dietary studies, sampling at a single consistent time on each of these days ~~would be the minimum, but collection of data at three or ideally five to six~~is sufficient, whereas gavage dosing may warrant additional sampling times ~~on each of these days would support~~to obtain a better ~~estimation~~estimate of ~~dose metrics, such as area under the curve in blood and minimum or maximum~~the range of internal doses. However, it is not necessary to generate a full concentration ~~[2].~~time-course on any of the sampling days.. If necessary, blood can be pooled by sex within litters for fetal and neonatal analyses.

22. Dietary exposure is the preferred method of administration. If gavage studies are performed, it should be noted that the pups will normally only receive test substance indirectly through the milk, until direct dosing commences for them at weaning. In diet or drinking water studies, the pups will additionally receive test substance directly when they commence eating for themselves during the last week of the lactation period. Modifications to the study design should be considered when excretion of the test substance in milk is poor and where there is lack of evidence for a continuous exposure of the offspring. In these cases direct dosing of pups already during the lactation period should be considered based on toxicokinetic information, offspring toxicity or changes in bio-markers [3, 4]. Careful consideration of benefits and disadvantages should be made prior to conducting direct dosing studies on nursing pups [5].

Dosing schedule and administration of doses

23. Some information on oestrus cycles, male and female reproductive tract histopathology and testicular/epididymal sperm analysis will be available already from previous systemic toxicity studies of 28 day and/or 90 day duration. The duration of the pre-mating treatment in the Extended One-Generation Study therefore is aimed at the detection of effects on ~~libido and other~~ functional changes that may interfere with mating behaviour and fertilisation. The pre-mating treatment should be sufficiently long to achieve steady-state exposure conditions in P males and females. A 2-week pre-mating treatment for both sexes is considered adequate in most cases. For females, this covers 3-4 complete oestrus cycles and should be sufficient to detect any adverse effects on cyclicity. For males, this is equivalent to the time required for epididymal transit of maturing spermatozoa and should allow the detection of post-testicular effects on sperm (during the final stages of spermiation and epididymal sperm maturation) at mating. Testicular and

[7] ~~**NOTE: Neonate blood samples**~~

~~Additional information on milk ingestion (time since last nursing, amount ingested) may be required to evaluate neonate blood concentrations of test substance or metabolite(s).~~

epididymal histopathology and analysis of sperm parameters is scheduled at termination of the P and F_1 males after exposure for at least the time required for one complete course of spermatogenesis [8].

24. The main objective of the Extended One-Generation Study is to evaluate specific life stages not covered by other types of toxicity studies (*e.g.* mating, pregnancy, parturition, lactation, offspring development prenatal to adult). Pre-mating exposure scenarios for males could be adapted, therefore, if testicular toxicity (impairment of spermatogenesis) or effects on sperm integrity and function have been clearly identified in previous studies. Similarly, for females, known effects of the test substance on the oestrous cycle and thus sexual receptivity, may justify different pre-mating exposure scenarios. In special cases it may be acceptable that treatment of the P females is initiated only after a sperm-positive smear has been obtained [9].

25. Once the pre-mating dosing period is established, the animals should be treated with the test substance continuously on a 7-days/week basis, from an age of about 11 weeks until necropsy. All animals should be dosed by the same method. Dosing should continue during the 2-week mating period and, for P females, throughout gestation and lactation up to the day of weaning. Males should be treated in the same manner until termination at the time when the F_1 animals are weaned. For necropsy, priority should be given to females; they should be necropsied on the same/similar day of lactation. Necropsy of males can be spread over a larger number of days, depending on laboratory facilities. Unless already initiated during the lactation period, direct dosing of the selected F_1 males and females should begin at weaning and continue until scheduled necropsy ~~on PND 70 or 90~~, depending on cohort assignment (see Figure 1).

[8] **NOTE: Premating exposure duration and spermatogenesis**
In adult testes all germ cell stages are present simultaneously. The action of a testicular toxicant on sensitive cell populations may therefore be detected by histopathology at a very early stage after the effect has been elicited, in many cases just a few days after the administration of the test compound. Collaborative studies and review papers have shown that for rodents a direct evaluation of testicular changes, conducted 2-4 weeks after initiation of dosing, reliably detects effects on spermatogenesis and is more sensitive than a mating test [6, 7, 8, 9]. Because of the highly efficient process of sperm production in these animal species a severe decrease in sperm output is necessary to observe a reduction in male fertility and less obvious reductions may often go undetected in mating tests, even when the pre-mating exposure of males is extended to 10 weeks. Good testicular histopathology will detect also the more subtle effects, e.g. the partial loss of germ cell stages, at doses that do not yet prevent the male from siring litters, and may also allow an identification of the most sensitive cell population within the testis [10]. This TG, therefore, relies on the most sensitive method to examine effects on spermatogenesis, i.e. histopathological examination of the testes combined with the analysis of sperm counts, motility, and abnormalities. If P males are evaluated at weaning of the F1 generation, 8-10 weeks after initiation of dosing and F1 males at the age of 90 days, the total treatment period covers at least 4 cycles of the spermatogenic epithelium and will allow for cumulation and propagation of effects in the testis and in the epididymis. In addition, F1 weanlings may be examined for the appropriate onset of spermatogenesis.

[9] **NOTE: Ensuring the production of an F_1 generation when testing reproductive toxicants**

The duration of the pre-mating treatment in males might be adjusted if the sensitive germ cell stage is known, to focus on the evaluation of maternal and offspring endpoints. Generally, premeiotic and meiotic spermatocytes are considered the most vulnerable cell populations. Any effect elicited at these stages would require 3-4 weeks to become visible as a reduction in epididymal sperm numbers so that a 2-week pre-mating treatment would be compatible with these males being still fully fertile during the subsequent 2-week mating period. However, effects elicited on elongating/elongated spermatids may require a reduction in the duration of the pre-mating treatment in order to obtain litters at the affected dose levels.

If the test substance is known to impair sperm function (e.g. decrease sperm motility) to an extent that is clearly incompatible with pregnancy induction it may be advisable to pair untreated males with treated females at the respective dose levels in order to be able to evaluate effects on the dams and the F_1.

26. For substances administered via the diet or drinking water, it is important to ensure that the quantities of the test substance involved do not interfere with normal nutrition or water balance. When the test substance is administered in the diet either a constant dietary concentration (ppm) or a constant dose level in terms of the body weight of the animal may be employed; the option chosen must be specified.

27. When the test substance is administered by gavage, ~~this should be done using a stomach tube, or other specific gavage instruments for this procedure. The~~the volume of liquid administered at one time should not exceed 1 mL/100 g body weight (0.4 mL/100 g body weight is the maximum for oil, e.g. corn oil). A volume of up to 2 mL/100 g body weight may be used for aqueous solutions in exceptional cases. Except for irritant or corrosive substances which will normally reveal exacerbated effects with higher concentrations, variability in test volume should be minimised by adjusting the concentration to ensure a constant volume at all dose levels. The treatment should be given at similar times each day. The dose to each animal should normally be based on the most recent individual body weight determination and adjusted at least weekly in adult males and adult non-pregnant females, and every two days in pregnant females and animals during the 2 weeks following weaning, to maintain a constant dose level in terms of animal body weight. However, should TK data indicate a low placental transfer of the test substance, the gavage dose during the last week of pregnancy may have to be adjusted to prevent administration of an excessively toxic dose to the dam. Females should not be treated by gavage, or any other route of treatment where the animal needs to be handled, on the day of parturition; omission of test substance administration on that day is preferable to a disturbance of the birth process.

Mating

28. Each P female should be placed with a single, randomly selected male from the same dose group (1:1 mating) until evidence of copulation is observed or either 3 estrous periods or 2 weeks have elapsed. Where one or a few males have died, a 1M:2F mating should be permitted such that all females are paired. The day on which a vaginal plug or sperm is found is designated as Day 0 of pregnancy. Animals should be separated as soon as possible after evidence of copulation is observed. If mating has not occurred after 2 weeks or 3 estrous periods, the animals should be separated without further opportunity for mating. Mating pairs have to be clearly identified in the data.

Litter size

~~*[original paragraph]*~~

29. Standardization of litter size to 10 is recommended in order to remove the confounder of litter size on survival, growth, body weights, acquisition of developmental landmarks, etc. If standardization is performed, the following procedure should be used. On day 4 after birth, the size of each litter may be adjusted by eliminating extra pups by random selection to yield, as nearly as possible, five males and five females per litter. Selective elimination of pups, *i.e*. based upon body weight, is not appropriate. Whenever the number of male or female pups prevents having five of each sex per litter, partial adjustment (for example, six males and four females) is acceptable. Adjustments are not appropriate for litters of ten pups or less. The surplus pups are subject to gross necropsy with detailed visceral examination for the detection of possible developmental abnormalities and consideration given to measuring serum thyroid hormone concentrations. If necessary, neonatal (PND 4) blood can be pooled by sex within litters for biochemical~~/~~/thyroid hormone analyses.

~~#~~

~~*[Alternative]*~~

~~29. Pregnant females are allowed to litter normally and rear all their offspring to weaning. This will allow to detect prenatal effects on the offspring that become manifest during the lactation period and maximise the number of F_1 animals available for distribution into cohorts on PND 21. In addition, the full reproductive capacity of the P females is established at each dose level. Reduction of litter size to a standard number of pups is not recommended as it will result in loss of information on postnatal manifestations of developmental defects without removing the confounding of the original litter size on growth and attainment of developmental landmarks [10].~~

Selection of pups for post-weaning studies

30. At weaning (around PND 21), at least 5 male and 5 female pups (if possible) from each litter are selected for further examinations. Pups are selected randomly with the exception that obvious runts (animals with a body weight more than two standard deviations below the mean pup weight of the respective litter) should not be included as they are unlikely to be representative of the treatment group~~.~~. All selected pups are kept at least until puberty and examined for sex-specific maturational endpoints (preputial gland separation, vaginal opening) regardless of cohort assignment.

The remaining pups, including runts, are subject to gross necropsy. Specified organs are weighed and preserved for possible histopathological examinations. Serum thyroid hormones (T4 and TSH) are measured. Alternatively, if the compound has suspected neurotoxic effects, animals culled at weaning may be used for an additional detailed examination of neurotoxicity/neuropathology on PND 21.

31. On PND21, the selected F_1 pups are assigned to one of three cohorts of animals as follows:

Cohort 1 = Reproductive/developmental toxicity
Cohort 2 = Developmental neurotoxicity/clinical endpoints
Cohort 3 = Developmental immunotoxicity

32. At least 1 pup per sex per litter (if possible) is assigned to each cohort by random selection. ~~To assure an equal distribution of littermates and avoid body weight disparities, animals are assigned to each cohort/dose group by using a body weight stratified design (similar to the procedure used for random distribution of P animals to dose groups before initiation of treatment).~~ Should there be an insufficient number of pups in a litter to serve all cohorts, Cohort 1 takes precedence as it can be extended if necessary to produce an F_2 generation. Treatment groups where less than 10 litters survive to weaning may be discontinued. More than 1 pup per sex per litter may be assigned to any of the cohorts in case of specific concern, i.e. if a chemical is suspected to be either a neurotoxicant, immunotoxicant or reproductive toxicant. The additional

~~[10] **Note: Litter size and development**~~

~~Differences in pup body weight at weaning are determined to a great extent by prenatal factors which influence birth weight, such as number of implantations, embryonal loss before the foetal phase, duration of pregnancy and litter size at birth. Postnatal factors relate to nursing ability of the dam (milk volume, energy expenditure of the litter), fat/energy content of the milk, and finally the number of pups that compete for these resources. Except for the last factor, none of these can be standardised by litter size reduction. Removing pups on PND 4 to obtain a standard number may equalize the care dams lavish on their pups in each litter, an important factor in neurodevelopment of the pups [11] but similar stimulation may be obtained by daily handling of the pups. Offspring from larger litters will have significantly more milk available after culling than pups from smaller litters, and the resulting growth spurt in a subset of litters may increase variability. In addition, an increased neonatal access to food has been shown to enhance the development of obesity and the metabolic syndrome in offspring, including diet-induced alterations in brain development [12]. Litter size is not associated with survival during the lactation period, except at the low end of the distribution [13, 14].~~

pups may be used for examinations at different timepoints, for the evaluation of supplementary endpoints, or to increase sample size in the cohort. However, for the analysis of the data, it must be kept in mind that the basic statistical unit is the litter, not the individual pup.

33. The animals not allocated to any of the cohorts will be terminated after puberty has been achieved, unless the results indicate the need for further in-life investigations in an extended number of F_1 subjects. Gross necropsy is performed and specified organs weighed and preserved for possible histopathological examinations. Serum thyroid hormones (T4 and TSH) are measured.

Second mating of the P animals

34. A second mating of P animals might only be conducted when an equivocal effect results from the first mating. This is not normally recommended for the P females as it comes at the expense of losing important information on the number of implantation sites (and thus post-implantation and peri-natal loss data, indicators of a possible teratogenic potential) for the first litter. The need to verify or elucidate an effect in exposed females, would be served better by extending the study to include a mating of the F_1 generation. However, a second mating of the P males with untreated females is always an option to clarify equivocal findings or for further characterisation of effects on fertility observed in the first mating.

IN-LIFE OBSERVATIONS

Clinical observations

35. For the P and the selected F_1 animals a general clinical observation is made each day. In the case of gavage dosing its timing should ~~take into account the anticipated peak period of effects.~~be prior to dosing. Pertinent behavioural changes, signs of difficult or prolonged parturition (for P females) and all signs of toxicity are recorded. Twice daily, during the weekend once daily, all animals are observed for morbidity and mortality.

36. In addition, a more detailed examination of ~~each P animal~~all P and F1 animals (after weaning) is conducted on a weekly basis and could conveniently be performed on an occasion when the animal is weighed~~, which would minimize handling stress.~~. Observations should be carefully conducted and recorded~~, preferably by observers unaware of the treatment group assignment of the animals and~~ using scoring systems that have been defined by the testing laboratory. Efforts should be made to ensure that variations in the test conditions are minimal. Signs noted should include, but not be limited to, changes in skin, fur, eyes, mucous membranes, occurrence of secretions and excretions and autonomic activity (*e.g.*, lacrimation, piloerection, pupil size, unusual respiratory pattern). Changes in gait, posture, response to handling as well as the presence of clonic or tonic movements, stereotypy (*e.g.* excessive grooming, repetitive circling) or bizarre behaviour (*e.g.* self-mutilation, walking backwards) should also be recorded.

Body weight and food/water consumption

37. P animals are weighed on the first day of dosing and at least weekly thereafter ~~for adult males and adult non-pregnant females, and every two days in pregnant females (see paragraph 27).~~. In addition, P females are weighed during lactation on the same days as the weighing of the pups in their litters. All F_1 animals are weighed individually at weaning (PND 21)~~The selected F_1 animals are weighed every two days during the 2 weeks following weaning~~ and at least weekly thereafter. Body weight is also recorded on the day when they attain puberty (completion of preputial

separation or vaginal patency~~, between PND30 and PND50).~~). All animals are weighed at sacrifice.

38. During the study, food and water consumption (in the case of test substance administration in the drinking water) is recorded at least weekly on the same days as animal body weights (except during cohabitation).. The food consumption of each cage of F_1 animals is recorded weekly commencing from selection.

Clinical biochemistry / Haematology

39. To determine potential maternal toxicity, pregnant and lactating P females are monitored for standard indicators of toxicity and key markers identified from clinical signs, haematological, clinical chemistry or urinalysis assessments in previous repeat-dose toxicity studies. Ideally, the assessments should be conducted at the estimated time of maximum sensitivity (considerating TK and/or toxicodynamic data), if this is known. The procedures should be conducted with as little disturbance as possible to normal function; if this cannot be achieved, consideration should be given to conducting such assessments in a satellite group of animals.

40. When monitoring for systemic effects, blood samples from a defined site are taken on ten randomly selected P males and females per dose group at termination, stored under appropriate conditions and subjected to partial or full scale haematology[11], clinical biochemistry[12] or other examinations suggested by the known effect profile of the test substance[~~13~~ 14]. Animals are fasted prior to taking samples. In addition, blood from all animals may be taken and stored for possible later analysis to help clarify equivocal effects or to generate internal exposure data. If a second mating of P animals is not intended, the blood samples are obtained just prior to or as part of the procedure at scheduled sacrifice. In the case animals are retained, blood samples should be collected a few days before the animals are mated for the second time. Urinalysis can be

[11] **NOTE: Haematology parameters**

The following haematological parameters should be examined: haematocrit, haemoglobin concentration, erythrocyte count, total and differential leukocyte count, platelet count and blood clotting time/potential.

[12] **NOTE: Clinical chemistry parameters**

Investigations of plasma or serum should include: glucose, total cholesterol, urea, creatinine, total protein, albumin, at least two enzymes indicative of hepatocellular effects (such as alanine aminotranferase, aspartate aminotransferase, alkaline phosphatase, gamma glutamyl transpeptidase and sorbitol dehydrogenase). Measurements of additional enzymes and bile acids may provide useful information under certain circumstances.

~~[13] NOTE: Other parameters~~

~~Serum markers of acute tissue damage should be considered for chemicals in certain classes or on a case-by-case basis. If a specific potential toxic effect of the test substance has been observed using special techniques on repeated-dose studies, these should also be used in this study (e.g. cholinesterase activity in plasma, red blood cells, brain and peripheral nervous tissue for compounds known to inhibit these enzymes; blood methaemoglobin concentration for compounds known to increase methaemoglobin formation; specific hormone measurements for endocrine modulators).~~

[14] **NOTE: Other parameters**

Serum markers of acute tissue damage should be considered for chemicals in certain classes or on a case-by-case basis. If a specific potential toxic effect of the test substance has been observed using special techniques on repeated-dose studies, these should also be used in this study (e.g. cholinesterase activity in plasma, red blood cells, brain and peripheral nervous tissue for compounds known to inhibit these enzymes; blood methaemoglobin concentration for compounds known to increase methaemoglobin formation; specific hormone measurements for endocrine modulators).

performed any time during the study period as long as the procedure does not interfere with the main objectives of the Extended One-Generation Study.[15]

41. For the investigation of pre- and postnatally induced major toxic (functional) effects on non-reproductive tissues, 10 male and 10 female Cohort 1 animals from each treatment group (1 male or 1 female per litter; all litters represented by at least 1 pup; randomly selected) will be subject, at termination, to standard clinical chemistry, including the assessment of ~~plasma~~serum levels for thyroid hormones (T4 and TSH), haematology including differential cell counting, analysis of lymphocyte subpopulations, including enumeration of natural killer cells~~,~~ in the spleen, bone marrow cellularity, and urinalysis assessments.. In case of an insufficient number of litters or pups in any dose group or when the study includes only one cohort, animals from the reproductive toxicity cohort may be evaluated also for the clinical biochemistry endpoints. At termination, spleens and adrenals will also be weighed, and one half of the spleen will be used for analysis of lymphocyte subpopulations, including enumeration of natural killer cells (see paragraph 68).

Oestrus cycles

42. Preliminary information of test substance-related effects on oestrus cycle may already be available from previous repeat-dose toxicity studies and may be used in designing a test substance-specific protocol for the Extended One-Generation Study. If such data are lacking, the assessment of vaginal cytology ~~perfomed to avoid inclusion of non-cycling P females into the study is continued daily~~is performed from the ~~initiation~~onset of the treatment period until confirmation of mating or the end of the 2-week mating period. When obtaining vaginal/cervical cells, care should be taken to avoid disturbance of mucosa and subsequently, the induction of pseudopregnancy [15, 16].

43. ~~If existing data show that oestrus cycles are not affected by the test substance at the dose levels selected for the Extended One-Generation Study, an evaluation of the oestrus cycle in P females by vaginal smears prior to the mating period is optional.~~

44. Oestrous cycles should be monitored daily for all F_1 females in Cohort 1 ~~for at least two weeks~~ after the onset of vaginal patency until the first oestrus and then again for a period of two weeks commencing ~~on~~around PND 75 ~~and continuing until PND 90 (the time of the F_1 mating if that were to occur).~~. Should a mating of the F_1 generation be necessary, the vaginal cytology would be followed until finding a sperm-positive smear.

45. For P and F_1 females the oestrus stage at termination is determined to allow correlation with histo(patho)logic findings in reproductive organs and hormone measurements, if applicable.

Mating and pregnancy

46. In addition to the standard endpoints (body weight, feed consumption, clinical observations including mortality/morbidity checks), the dates of pairing, the date of insemination and the date of delivery is recorded and the precoital interval (pairing to insemination) and the duration of pregnancy (insemination to delivery) are calculated. The P females should be examined carefully

[15] **NOTE: Urinalysis**
Unless existing data from repeated dose indicate that the parameter is not affected by the test substance, the following parameters should be evaluated: appearance, volume, osmolality or specific gravity, pH, protein, glucose, blood and blood cells, cell debris. Urine may also be collected to monitor excretion of test substance and/or metabolite(s).

at around the time of expected parturition for any signs of dystocia (difficult or prolonged parturition). Any abnormalities of nesting or nursing performance should be recorded.

47. The day on which parturition occurs is lactation day 0 (LD 0) for the dam and postnatal day 0 (PND 0) for the offspring. Alternatively, all comparisons may be based on post-coital time to eliminate confounding of postnatal development data by differences in the duration of pregnancy. This is especially important when the test substance exerts an influence on the duration of pregnancy.

Offspring parameters

48. Each litter should be examined as soon as possible after delivery (PND 0 or 1) to establish the number and sex of pups, stillbirths, live births, and the presence of gross anomalies (externally visible abnormalities, including cleft palate; subcutaneous haemorrhages; abnormal skin colour or texture; presence of umbilical cord; lack of milk in stomach; presence of dried secretions). In addition, the first clinical examination of the neonates should include a qualitative assessment of body temperature (normal or cool to touch), state of activity (normal, hyperactive, drowsy) and reaction to handling (normal, struggling, vocalisation, none).. Pups found dead on PND 0 or at a later time should be examined for possible defects and cause of death. Live pups are counted and weighed individually on PND 0 or PND 1, and regularly thereafter, e.g., at least on PND 4, 7, 14, and 21. Clinical examinations as applicable for the age of the animals should be repeated at the times when the offspring are weighed, or more often in case specific findings have been made at birth. Signs noted could include, but not be limited to, changes in skin, fur, eyes, mucous membranes, occurrence of secretions and excretions and autonomic activity. Changes in gait, posture, response to handling as well as the presence of clonic or tonic movements, stereotypy or bizarre behaviour should also be recorded.

49. The anogenital distance of each pup should be measured between PND 0 and PND 4 ~~PND 2, and if an effect is noted, also on PND 21 and at necropsy.~~. Pup body weight should be collected on the day the AGD is measured and the AGD should be normalized to a measure of pup size, preferably the cube root of body weight (49). Presence of nipples/areolae in male pups should be checked on PND 12 or 13.

50. Physical or behavioural abnormalities observed in the dams or offspring should be recorded. Physical development of the offspring should be followed mainly by body weight gain. ~~Other~~Evaluation of other physical parameters (*e.g.* pinna detachment, auditory canal and eye opening, incisor eruption, hair growth, surface righting reflex, attainment of hearing ability/auditory startle response) may be conducted as they may give supplementary information and serve as markers for endocrine effects and nutritional deficiencies. ~~Measurements for these endpoints should be conducted either from before beginning of attainment until completion (taking care to handle all pups on all days in the same way) or on a single day within the control range of achievement which allows the detection of acceleration as well as delay in the development of the respective endpoint [16].~~

~~[16] **Note: Evaluation of pre-weaning developmental landmarks**~~

~~Attainment of landmarks is dependent on body weight and gestational age at birth in many cases but has also been shown to be related to hormonal status of the pups. Auditory canal opening, attainment of hearing ability/auditory startle response, incisor eruption, eye opening, and hair growth may be accelerated or delayed by the availability of thyroid hormones [17, 18, 19, 20]; hair growth and hair growth patterns can be affected also by glucocorticoids, insulin-like growth factor I, or prolactin [20]; impairment/delay of surface righting reflex has been associated with iron deficiency and defects in myelination [21].~~

51. ~~All selected~~At least three F_1 females ~~(Cohorts 1, 2 and 3)~~per litter are evaluated daily for vaginal patency commencing before the expected day of vaginal patency, to allow detection of early occurrence ~~of earlier occurrence.~~. Any abnormalities such as a vaginal thread should be noted. All selected F_1 males (Cohorts 1, 2 and 3) are evaluated daily for balano-preputial separation commencing ~~from PND 35.~~before the expected day of balano-preputial separation. Sexual maturity of F_1 females and males is compared to physical development by determining age and body weight at vaginal opening or balano-preputial separation, respectively [22].

Assessment of potential developmental neurotoxicity (Cohort 2~~)~~)

52. At least 10 male and 10 female Cohort 2 animals from each treatment group (1 male or 1 female per litter; all litters represented by at least 1 pup; randomly selected) should be used for assessment of neurotoxicity, including neurohistopathology. The same animals ~~may~~should be subjected to functional observation battery, motor activity and neuropathology assessments. Efforts should be made to ensure that variations in all test conditions are minimal and are not systematically related to treatment. Among the variables that can affect behaviour are sound level, temperature, humidity, lighting, odors, time of day, and environmental distractions. All animals should be observed carefully by trained observers who are unaware of the animals' treatment status, using standardized procedures to minimize observer variability. Where possible, it is advisable that the same observer evaluates the animals in a given test. If this is not possible, some demonstration of inter-observer reliability is required. For each parameter in the behavioural testing battery explicit, operationally defined scales and scoring criteria are to be used. If possible, objective quantitative measures should be developed for observational endpoints which involve subjective ranking.

53. At an appropriate time between PND 49 and PND 56 the F_1 animals are subjected to a functional observational battery that includes a thorough description of the subject's appearance, behavior, and functional integrity. This is assessed through observations in the home cage, after removal to a standard arena for observation (open field) where the animal is moving freely, and through manipulative tests. Testing should proceed from the least to the most interactive. A list of measures is presented in Appendix A.

54. *## alternative wording to be inserted ###* Motor activity should be monitored at least once between PND 63-70. However, if the compound undergoing testing is known to have neurotoxic potential, then motor activity may be assessed at earlier ages (*e.g.,* PND 13, 17, and 21) in the same animals. Each animal is tested individually. The test session should be long enough to demonstrate intra-session habituation for non-treated controls. Motor activity should be monitored by an automated activity recording apparatus which should be capable of detecting both increases and decreases in activity, (*i.e.,* baseline activity as measured by the device should not be so low as to preclude detection of decreases, nor so high as to preclude detection of increases in activity). Each device should be tested by standard procedures to ensure, to the extent possible, reliability of operation across devices and across days. To the extent possible, treatment groups should be balanced across devices. Treatment groups should be counter-balanced across test times to avoid confounding by circadian rhythms of activity.

55. If existing information indicates the need for other functional testing (*e.g.*, sensory, social, cognitive), these should be integrated without compromising the integrity of the other evaluations conducted in the study. If this testing is performed in the same animals as used for standard FOB and motor activity testing, different tests should be scheduled to minimise the risk of compromising the integrity of these tests. Supplemental procedures may be particularly useful

when empirical observation, anticipated effects, or mechanistic/mode-of-action indicate a specific type of neurotoxicity.

Assessment of developmental immunotoxicity (Cohort 3)

56. At an appropriate time between PND ~~50~~42-70, 10 male and 10 female Cohort 3 animals from ~~at least 10 litters per~~ each treatment group (1 male or 1 female per litter; all litters represented by at least 1 pup; randomly selected) should be used to assess the primary IgM antibody response to a T cell dependent antigen, such as Keyhole Limpet Hemocyanin (KLH) or Sheep Red Blood Cells (SRBC), consistent with current immunotoxicity testing procedures~~.~~ (53)(54). The response may be evaluated by counting specific plaque forming cells (PFC) in the spleen or by determining the titer of specific antibody in the serum by ELISA, at the peak of the response. Responses typically peak four (PFC response) or five (ELISA) days after intravenous immunization, although this should be confirmed by each laboratory before testing begins. Exposure to the test article will continue until the day before collecting spleens for the PFC response or serum for the ELISA assay.

~~If the antibody response is similar in all groups, is marginally affected or data (e.g., lymphoid organ weights or histopathology) from range-finding or other preliminary studies suggest possible immune system effects, correlates of cell mediated immune function may be evaluated in the remaining 10 male and 10 female pups in Cohort 3. Humoral immune function data can be had within a few days after sacrifice and the remaining Cohort 3 pups should therefore remain on exposure until additional testing is made. Cell mediated immune function assays include the delayed type hypersensitivity response, cell mediated cytotoxicity, or the natural killer cell assay (preferably stimulated by poly I:C, if testing is conducted in mice). Exposure to the test article should continue though the day before sacrifice.~~

Assessment of potential reproductive toxicity (*Triggered*)

57. Cohort 1 animals can be maintained on treatment beyond PND 90 and bred to obtain a F_2 generation if data from previous studies or early findings from this study raise concern for possible effects on reproduction or if equivocal effects on reproduction are obtained in the P animals~~.~~. Males and females of the same dose group should be cohabited (avoiding the pairing of siblings) for up to two weeks, beginning on approximately PND 90. Procedures should be similar to those for the P animals, although it may suffice to terminate the litters on PND 4 rather than follow them to weaning or beyond, depending on the concerns that triggered the extension of the study. The decision of whether a breeding of the F_1 generation is necessary should be based on a weight of the evidence approach that considers the nature and degree of the effects found in the F_1 generation, as well as other available pertinent information (*e.g.*, mechanism studies, subchronic toxicity studies). A more detailed discussion of triggers is presented in Appendix B.

TERMINAL OBSERVATIONS

Sperm parameters

58. Sperm parameters may have already been examined as part of ~~a 28- or 90-day~~ systemic toxicity ~~study.~~studies. Such data may be helpful in the planning of the Extended One-Generation Study as well as in the interpretation of the results. If the existing data show that sperm parameters are not affected by the test substance and if the mating of the P generation does not produce evidence for impaired sperm function the analysis of sperm parameters is optional for P males. However, in the case that clear or equivocal effects on reproduction (*e.g.* reductions in sperm-positive or fertile matings, decreased litter size, histopathologic evidence of impaired

spermatogenesis from previous repeat-dose studies) are observed, a quantitation of epididymal sperm and the analysis of functional and morphological sperm parameters may help to confirm or characterise the reproductive toxicity.

59. At termination, testis and epididymis weights are recorded for all P and F_1 males (Cohort 1). At least one testis and one epididymis are reserved for histopathological examination. The remaining epididymis is used for enumeration of cauda epididymis sperm reserves [23, 24]. In addition, sperm from the cauda epididymis (or vas deferens) is collected using methods that minimise damage for evaluation of sperm motility and morphology. [25]. One testis is reserved for the evaluation of absolute and relative amounts of the different germ cell stages, by using preferably cell sorting procedures. Enumeration of homogenization-resistant spermatids may be used in principle, but would be considered less informative.

60. Sperm motility can either be evaluated immediately after sacrifice or recorded for later analysis. The percentage of progressively motile sperm could be determined either subjectively or objectively by computer-assisted motion analysis [26, 27, 28, 29, 30, 31]. For the evaluation of sperm morphology an epididymal (or vas deferens) sperm sample should be examined as fixed or wet preparations [32] and at least 200 spermatozoa per sample classified as either normal (both head and midpiece/tail appear normal) or abnormal. Examples of morphologic sperm abnormalities would include fusion, isolated heads, and misshapen heads and/or tails (51). Misshapen or large sperm heads may indicate defects in spermiation.

61. If sperm samples are frozen, smears fixed and images for sperm motility analysis recorded at the time of necropsy [33], subsequent analysis may be restricted to control and high-dose males unless treatment-related effects are observed; in that case, the lower dose groups should also be evaluated.

Gross necropsy

62. At the time of termination or death during the study, all P and F_1 animals [17] are weighed and subjected to gross necropsy, being examined macroscopically for any structural abnormalities or pathological changes. This includes careful examination of the external surface of the body, all orifices, and the cranial, thoracic and abdominal cavities and their contents. Special attention should be paid to the organs of the reproductive system. Pups that are humanely killed in a moribund condition and dead pups should be recorded and, when not macerated, examined for possible defects and/or cause of death and preserved.

63. For adult P and F_1 females, a vaginal smear is examined on the day of necropsy to determine the stage of the estrous cycle. The uteri of all cohabited P females (and F_1 females, if applicable) are examined for the presence and number of implantation sites, in a manner which does not compromise histopathological evaluation.

Organ weights – P animals

64. At the time of termination, body weight and the wet weight of the following organs of all P animals are determined (paired organs should be weighed individually or paired) as soon as possible after dissection to avoid drying:

- Uterus (with oviducts and cervix), paired ovaries

[17] **NOTE: F1 gross necropsy**

Includes all pups removed during the lactation phase, the pups not selected for continuation at weaning and offspring terminated at sexual maturation as well as all offspring continued to adult age.

- Testes, epididymides (total and cauda for either one or both)
- Prostate (dorsolateral and ventral part combined) + seminal vesicles with coagulating glands and their fluids (as one unit). Care should be exercised when trimming the prostate complex to avoid puncture of the fluid filled seminal vesicles

Brain, liver, kidneys, heart, lung, spleen, thymus, pituitary, adrenal glands and known target organs or tissues

The following organ weights should be determined also after fixation:

- Thyroid (trimming should also be done very carefully and only after fixation to avoid tissue damage. Tissue damage could compromise histopathology analysis)
- Pituitary
- Dorsolateral and ventral part of the prostate separately after separation

Samples of peripheral nerve, muscle and spinal cord should be preserved for histopathological examination. Vas deferens in males should also be preserved for histopathological examination.

Organ weights and tissues retained – F1 animals (### subject to rephrasing ###)

65. From pups subject to gross necropsy on PND 21 and from F_1 animals not continued after attainment of puberty the following organs are weighed (where appropriate) and retained in an appropriate fixative:

- Ovaries, uterus (with oviducts and cervix), mammary gland in the female
- Testes, epididymides, vas deferens, seminal vesicles with coagulating glands and their fluids (as one unit), prostate and mammary gland in the male
- Brain, pituitary, eye plus optic nerve, liver, GI tract, kidneys, urinary bladder, heart, lung, trachea (weigh post-fixation, pre- fixation it will include the thyroid), thyroid (weigh post-fixation), spleen, thymus, –adrenal glands, bone marrow, known target organs or tissues

~~**Cohort 2 animals**~~

~~66. In addition to the organs listed for the P animals the following tissues from the Cohort 2 animals subject to developmental neurotoxicity assessments are weighed after perfusion fixation:~~

- ~~Eye plus optic nerve~~

~~The remaining Cohort 2 animals should be subject to gross necropsy after blood and urine samples have been taken and evaluated only for abnormalities; which, if found, should be retained.~~

~~**Cohort 3 animals**~~

~~67. For Cohort 3 animals at gross necropsy the following tissues are weighed and fixed in addition to the organs listed above for the P animals:~~

~~An inguinal and a mesenteric lymph node~~ [18]

66.

67.

[19]

[18] ~~**Note: Lymph nodes**~~

~~All mesenteric lymph nodes present in the chain should be removed and weighed in aggregate, because individual nodes vary in size and weight.~~

[19] **Note: Lymph nodes**

Cohort 1 animals

68. Cohort 1 animals will be subject to gross necropsy either at PND 90 or, if a reproductive toxicity assessments of the F_1 is triggered, when they are no longer needed for assessment. The same organs as listed for the P animals are weighed and fixed for all Cohort 1 animals. After weighing, the spleen should be divided in half (by length), one half preserved for histopathology, and the other half processed for ~~flow cytometry (see paragraph 41).~~lymphocyte subset analysis (see paragraph 41), and lymph nodes in and distant from the route of exposure should be weighed.

Histopathology – P animals

69. Full histopathology of the organs listed in Paragraph 64 is performed for 10 randomly chosen high dose and control P animals per sex. Organs demonstrating treatment-related changes should also be examined in the remainder of the high-dose and control animals and for all animals at the lower dose groups to aid in determining a NOAEL. Additionally, reproductive organs of all animals suspected of reduced fertility, *e.g.*, those that failed to mate, conceive, sire, or deliver healthy offspring, or for which oestrous cyclicity or sperm number, motility, or morphology were affected, and all gross lesions should be subjected to histopathological evaluation. Multiple sections are examined from the brain to allow examination of olfactory bulbs, cerebral cortex, hippocampus, basal ganglia, thalamus, hypothalamus, mid-brain (tectum, tegmentum, cerebral peduncles) brain-stem, and cerebellum.

70. The postlactational ovary of the P females should contain primordial and growing follicles as well as the large corpora lutea of lactation. Histopathological examination should be aimed at detecting qualitative depletion of the primordial follicle population. An enumeration of follicular stages could assist in determining the cause of a dose-related decrease in litter size.

71. Besides examining gross lesions such as atrophy or tumors, detailed testicular histopathology examinations are conducted on P males in order to identify treatment-related effects such as retained spermatids, missing germ cell layers or types, multinucleated giant cells or sloughing of spermatogenic cells into the lumen [34]. Examination of the intact epididymis should include the caput, corpus, and cauda, which can be accomplished by evaluation of a longitudinal section [10]. The epididymis should be evaluated for leukocyte infiltration, change in prevalence of cell types, aberrant cell types, phagocytosis of sperm, and the absence of clear cells in the caudal epithelium [52].

Histopathology – F1 animals

Cohort 1 animals

72. Full histopathology of the organs listed in Paragraph 64 is performed for at least 10 randomly chosen high dose and control adult Cohort 1 animals per sex. All litters should be represented by at least 1 pup. Organs and tissues demonstrating treatment-related changes and all gross lesions should also be examined in the remainder of the high-dose and control animals and for all animals at the lower dose groups to aid in determining a NOAEL.

73. A ~~quantitative~~qualitative evaluation of primordial follicles should be conducted in the F_1 females; the number of animals, ovarian section selection, and section sample size should be

All mesenteric lymph nodes present in the chain should be removed and weighed in aggregate, because individual nodes vary in size and weight.

statistically appropriate for the evaluation procedure used. ~~Examination should include enumeration of the number of pre-antral and antral follicles, as well as primary follicles, which can be combined with small growing follicles, for comparison of treated and control ovaries [35, 36, 37, 38, 39, 40, 41, 42]. In case a reduction in the number of primordial follicles is observed ovaries from immature (PND 21) and pubertal females should be examined as well.~~ Histopathological examination should be aimed at detecting qualitative depletion of the primordial follicle population. Oviduct, uterus and vagina are examined for appropriate organ-typic development.

74. Detailed testicular histopathology examinations are conducted on the F_1 males in order to identify treatment-related effects on testis differentiation and development and on spermatogenesis. Sections examined should include the rete testis. Caput, corpus, and cauda of the epididymis and the vas deferens are examined for appropriate organ-typic development as well as for the parameters required for the P males.

Cohort 2

75. Neurohistopathology is performed for at least 10 high dose and control Cohort 2 animals per sex ~~on~~up to PND 90. Organs or tissues demonstrating treatment-related changes should also be examined for all animals at the lower dose groups to aid in determining a NOAEL. Multiple sections are examined from the brain to allow examination of olfactory bulbs, cerebral cortex, hippocampus, basal ganglia, thalamus, hypothalamus, mid-brain (thecum, tegmentum, cerebral peduncles), brain-stem and cerebellum. Additionally, the eyes (retina and optic nerve) and samples of peripheral nerve, muscle and spinal cord are examined. ~~In addition to standard staining techniques, separate sections of neural tissue are immunostained for glial fibrillary acidic protein, a marker for myelination (*e.g.*, myelin basic protein) and a synaptic marker (*e.g.*, synaptophysin).~~ Morphometric (quantitative) evaluations should be performed on representative areas of the central nervous system (homologous sections carefully selected based on reliable microscopic landmarks) and may include linear and/or areal measurements of specific brain regions. The neuropathologist should exercise appropriate judgment as to whether sections prepared for measurement are homologous with others in the sample set and therefore suitable for inclusion, since linear measurements in particular may change over a relatively short distance (50). Non-homologous sections should not be used. While the objective is to sample all animals reserved for this purpose (10/sex/dose level), smaller numbers may still be adequate. However, samples from fewer than 6 animals/sex/dose level would generally not be considered sufficient for the purposes of this test guideline. Stereology may be used to identify treatment-related effects on parameters such as volume or cell number for specific neuroanatomic regions. All aspects of the preparation of tissue samples, from the perfusion of animals, through the dissection of tissue samples, tissue processing, and staining of slides should employ a counterbalanced design such that each batch contains representative samples from each dose group. If morphometric or stereological analyses are to be used, then brain tissue must be embedded in appropriate media at all dose levels at the same time in order to avoid shrinkage artifacts known to be associated with prolonged storage in fixative. All neurohistological procedures should be consistent with OECD 426.

REPORTING

Data

76. Data are reported individually and summarised in tabular form, showing, where appropriate, for each test group and each generation the number of animals at the start of the test, the number of animals found dead during the test or killed for humane reasons, the time of any death or humane kill, the number of fertile animals, the number of pregnant females, the number of females giving birth to a litter, the number of animals showing signs of toxicity, a description of

the signs of toxicity observed, including time of onset, duration, and severity of any toxic effects, the types of histopathological changes, and all relevant litter data.

77. Numerical results should be evaluated by an appropriate, generally accepted statistical method; the statistical methods should be selected as part of the design of the study. The statistical methods should appropriately address non-normal data (*e.g.*, count data), censored data (*e.g.*, limited observation time such 2 min), non-independence (*e.g.*, litter effects and repeated measures), and unequal variances. Generalized linear mixed models cover a broad class of analytical tools that could be appropriate for the data generated under this TG–. Dose-response statistical models may be useful as well. The report should include sufficient information on the method of analysis and the computer program employed, so that an independent reviewer/statistician can re-evaluate and reconstruct the analysis.

Evaluation of results

78. The findings should be evaluated in terms of the observed effects including necropsy and microscopic findings. The evaluation includes the relationship, or lack thereof, between the dose of the test substance and the presence or absence, incidence and severity of abnormalities, including gross lesions, identified target organs, fertility, clinical abnormalities, reproductive and litter performance, body weight changes, mortality and any other toxic and developmental effects. The physico-chemical properties of the test substance, and when available, toxicokinetic data (including placental transfer and milk excretion) should be taken into consideration when evaluating test results.

79. ~~A properly conducted Extended One-Generation Reproduction Toxicity Study should provide a satisfactory estimation of a no-effect level and insight into possible adverse effects on fertiliy and reproduction, parturition, lactation and postnatal development including growth, sexual maturation and functional endpoints.~~

Test Report

80. The test report must include the following information:

Test substance:

- All available information on the substance, toxicokinetic and toxicodynamic properties of the test substance, available relevant information on structural analogues of the substance, pertinent results of previously conducted toxicity and toxicokinetic studies
- identification data
- purity

Vehicle (if appropriate):

- justification for choice of vehicle if other than water

Test animals:

- species/strain used
- number, age and sex of animals
- source, housing conditions, diet, nesting materials, etc.
- individual weights of animals at the start of the test
- vaginal smear data for P females before initiation of treatment
- P generation pairing records indicating male and female partner of a mating and mating success
- litter of origin records for adult F_1 generation animals

Test conditions:

- rationale for dose level selection
- details of test substance formulation/diet preparation, achieved concentrations
- stability and homogeneity of the preparation in the vehicle or carrier, in the blood and/or milk under the conditions of use and storage between uses
- details of the administration of the test substance
- conversion from diet/drinking water test substance concentration (ppm) to the achieved dose (mg/kg body weight/day), if applicable
- details of food and water quality (including diet composition, if available)

Results (summary and individual data):

- food consumption, water consumption if available, food efficiency (body weight gain per gram of food consumed), and test material consumption (for dietary/drinking water administration) for P and F_1 animals, except for the period of cohabitation and the last third of lactation
- absorption data (if available)
- body weight data for P animals
- body weight data for the selected F_1 animals postweaning
- time of death during the study or whether animals survived to termination
- nature, severity and duration of clinical observations (whether reversible or not)
- haematology, urinalysis and clinical chemistry data including TSH and T4
- toxic response data by sex and dose
- number of P and F_1 females with normal or abnormal oestrus cycle and cycle duration
- time to mating (precoital interval, the number of days between pairing and mating)
- toxic or other effects on reproduction, including numbers and percentages of animals that accomplished mating, pregnancy, parturition and lactation, of males inducing pregnancy, of females with signs of dystocia/prolonged or difficult parturition
- duration of pregnancy and, if available, parturition
- numbers of implantations, litter size and percentage of male pups
- number and percent of post-implantation loss, live births and stillbirths
- litter weight and pup weight data (males, females and combined), the number of runts if determined
- number of pups with grossly visible abnormalities
- toxic or other effects on offspring, postnatal growth, viability, etc.
- data on physical landmarks in pups and other postnatal developmental data
- data on sexual maturation of F_1 animals
- data on functional observations in pups and adults, as applicable
- body weight at sacrifice and absolute and relative organ weight data for the P and adult F_1 animals
- necropsy findings including organ weights
- detailed description of all histopathological findings
- total cauda epididymal sperm number, percent progressively motile sperm, percent morphologically normal sperm, and percent of sperm with each identified abnormality for P and F_1 males
- numbers and maturational stages of follicles contained in the ovaries of P and F_1 females, where applicable
- statistical treatment of results, where appropriate

add reporting of cohort 2 and 3 parameters
Include F2 parameters(as in OECD416)###

Discussion of results

Conclusions, including NOAEL values for parental and offspring effects

INTERPRETATION OF RESULTS

include cohort 2 and 3 spinoff

include references as appropriate

81. An Extended One-Generation Reproduction Toxicity Study will provide information on the effects of repeated exposure to a substance during all phases of the reproductive cycle. In particular, the study provides information on the reproductive parameters, and on development, growth, survival, and functional endpoints of offspring up to PND 90.

82. The results of the study should be interpreted taking into account all available information on the substance (i.e. physico-chemical (PC), toxicokinetic and toxicodynamic properties of the test substance, available relevant information on structural analogues of the substance (structure-activity relationships (SARs), results of previously conducted toxicity studies of the test substance (*e.g.* acute toxicity, toxicity after repeated application, mechanistic studies). Gross necropsy and organ weight results should be assessed in context with observations made in other repeat-dose studies, when feasible. Decreases in offspring growth might be considered in relationship to an influence of the test substance on milk composition [43].

83. ~~The results of this study can often be used in assessing the need for further testing of a chemical.~~ Extrapolation of the results of the study to man is valid to a limited degree and may be improved if comparative data on metabolism and mechanisms of toxicity can be incorporated. They are best used to provide information on no-effect-levels and permissible human exposure [44, 45, 46, 47].

LITERATURE

[1] Cooper, R. L., Lamb, J. C., Barlow, S. M., Bentley, K., Brady, A. M., Doerr, N., Eisenbrandt, D. L., Fenner-Crisp, P. A., Hines, R. N., Irvine, L. F. H., Kimmel, C. A., Koeter, H., Li, A. A., Makris, S. L., Sheets, L. P., Speijers, G. J. A., Whitby, K. E. (2006). A tiered approach to life stages testing for agricultural chemical safety assessment. Crit. Rev. Toxicol. 36, 69-98.

[2] Saghir, S.A., Mendrala, A.L., Bartels, M.J., Day, S.J., Hansen, S.C., Sushynski, J.M., Bus, J.S. (2006). Strategies to assess systemic exposure of chemicals in subchronic/chronic diet and drinking water studies. Toxicol Appl Pharmacol. 211, 245-260.

[3] Zoetis, T., Walls, I. (2003). Principles and Practices for Direct Dosing of Pre-Weaning Mammals in Toxicity Testing and Research. ILSI Press, Washington, DC.

[4] Moser, V. et al. (2005). Direct dosing of preweaning rodents in toxicity testing and research: Deliberations of an ILSI RSI expert working group. Int. J. Toxicol. 24, 87-94.

[5] Conolly, R.B. et al. (1999). Stimulating research to improve the scientific basis of risk assessment. Toxicol. Sci. 49, 1-4.

[6] Takayama, S. et al. (1995). A Collaborative Study in Japan on Optimal Treatment Period and Parameters for Detection of Male Fertility Disorders Induced by Drugs in Rats. J. Am. Coll. Toxicol. 14, 266-292.

[7] Ulbrich, B., Palmer, A. K. (1995). Detection of effects on male reproduction – a literature survey. J. Am. Coll. Toxicol. 14, 293-327.

[8] Mangelsdorf, I., Buschmann, J., Orthen, B. (2003). Some aspects relating to the evaluation of the effects of chemicals on male fertility. Regul. Toxicol. Pharmacol. 37, 356-369.

[9] Sakai, T. et al. (2000). Collaborative work to evaluate toxicity on male reproductive organs by repeated dose studies in rats - overview of the studies. Journal of Toxicological Sciences 25, 1-21.

[10] Creasy, D. M. (2003). Evaluation of testicular toxicology: A synopsis and discussion of the recommendations proposed by the Society of Toxicologic Pathology. Birth Defects Research (Part B) 68, 408-415.

[11] Weaver, I.C.G et al. (2006). Maternal care effects on the hippocampal transcriptome and anxiety-mediated behaviors in the offspring that are reversible in adulthood. Proc. Natl. Acad. Sci. USA 103, 3480–3485.

[12] Levin, B. E.(2006). Metabolic imprinting: critical impact of the perinatal environment on the regulation of energy homeostasis. Phil. Trans. R. Soc. B 361(1471), 1107-1121.

[13] Palmer, A. K., Ulbrich, B (1997). The cult of culling. Fundam. Appl. Toxicol. 38, 7-22.

[14] Agnish, N.D., Keller, K.A. (1997). The rationale for culling of rodent litters. Fundam. Appl. Toxicol. 38, 2-6.

[15] Goldman, J.M. et al., (2007). The rodent estrous cycle: Characterization of vaginal cytology and its utility in toxicological studies. Birth Defects Research Part B-Developmental and Reproductive Toxicology 80 (2), 84-97.

[16] Sadleir, R.M.F.S. (1979). Cycles and Seasons, In: *Reproduction in Mammals:* I. Germ Cells and Fertilization, C.R. Auston and R.V. Short (eds.), Cambridge, New York.

[17] Hoath, S.B., Pickens, W.L. (1987). Effect of thyroid hormone and epidermal growth factor on tactile hair development and craniofacial morphogenesis in the postnatal rat. Journal of Craniofacial Genetics and Developmental Biology 7 (2), 161-167.

[18] Freeman, S., Geal-Dor, M., Shimoni, Y., Sohmer, H. (1993). Thyroid hormone induces earlier onset of auditory function in neonatal rats. Hearing Research 69 (1-2), 229-235.

[19] Lakshmanan, J., Perheentupa, J., Alm, J., Fisher, D.A. (1986). Neonatal hyperthyroidism in mice had different effects on epidermal growth factor levels in submandibular gland, urine, and blood. Pediatric Research 20 (7), 628-631.

[20] Alonso, L.C., Rosenfield, R.L. (2003). Molecular genetic and endocrine mechanisms of hair growth. Hormone Research 60 (1), 1-13.

[21] Wu, L.-L., Zhang, L., Shao, J., Qin, Y.-F., Yang, R.-W., Zhao, Z.-Y. (2008). Effect of perinatal iron deficiency on myelination and associated behaviors in rat pups. Behavioural Brain Research 188 (2), 263-270.

[22] Korenbrot, C.C. et al., (1977). Preputial Separation as an External Sign of Pubertal development in the male rat. Biological Reproduction 17, 298-303.

[23] Gray, L.E. et al., (1989). A Dose-Response Analysis of Methoxychlor-Induced Alterations of Reproductive Development and Function in the Rat. Fundamental and Applied Toxicology 12, 92-108.

[24] Robb, G.W. et al., (1978). Daily Sperm Production and Epididymal Sperm Reserves of Pubertal and Adult Rats. Journal of Reproduction and Fertility 54, 103-107.

[25] Klinefelter, G.R. et al., (1991). The Method of Sperm Collection Significantly Influences Sperm Motion Parameters Following Ethane Dimethanesulfonate Administration in the Rat. Reproductive Toxicology 5, 39-44.

[26] Seed, J. et al. (1996). Methods for Assessing Sperm Motility, Morphology, and Counts in the Rat, Rabbit, and Dog: a Consensus Report. Reproductive Toxicology 10, 237- 244.

[27] Chapin, R.E. et al., (1992). Methods for Assessing Rat Sperm Motility. Reproductive Toxicology 6, 267-273.

[28] Klinefelter**,** G.R. et al., (1992). Direct Effects of Ethane Dimethanesulphonate on Epididymal Function in Adult Rats: an In Vitro Demonstration**.** Journal of Andrology 13, 409-421.

[29] Slott, V.L. et al., (1991). Rat Sperm Motility Analysis: Methodologic Considerations. Reproductive Toxicology 5, 449-458.

[30] Slott, V.L., Perreault, S.D., (1993). Computer-Assisted Sperm Analysis of Rodent Epididymal Sperm Motility Using the Hamilton-Thorn Motility Analyzer. In: Methods in Toxicology, Part A., Academic, Orlando, Florida. pp. 319-333.

[31] Toth, G.P. et al. (1989). The Automated Analysis of Rat Sperm Motility Following Subchronic Epichlorhydrin Administration: Methodologic and Statistical Considerations. Journal of Andrology 10, 401-415.

[32] Linder, R.E. et al., (1992). Endpoints of Spermatoxicity in the Rat After Short Duration Exposures to Fourteen Reproductive Toxicants. Reproductive Toxicology 6, 491-505.

[33] Working, P.K., Hurtt, M. (1987). Computerized Videomicrographic Analysis of Rat Sperm Motility. Journal of Andrology 8, 330-337.

[34] Russell, L.D. et al., (1990). Histological and Histopathological Evaluation of the Testis, Cache River Press, Clearwater, Florida.

[35] Heindel, J.J., Chapin, R.E. (eds.) (1993). Part B. Female Reproductive Systems, Methods in Toxicology, Academic, Orlando, Florida.

[36] Heindel, J.J. et al., (1989) Histological Assessment of Ovarian Follicle Number in Mice As a Screen of Ovarian Toxicity. In: Growth Factors and the Ovary, A.N. Hirshfield (ed.), Plenum, New York, pp. 421-426.

[37] Christian, M.S. (2007). Chapter 34: Test Methods for Assessing Female Reproductive and Developmental Toxicology in: *Principles and Methods of Toxicology* (5th Edition, A. Wallace Hayes, editor), pp. 1641-1711. Taylor & Francis: Boca Raton.

[38] Smith, B.J. et al,. (1991). Comparison of Random and Serial Sections in Assessment of Ovarian Toxicity. Reproductive Toxicology 5, 379-383.

[39] Heindel, J.J. (1999). Oocyte Quantitation and Ovarian Histology. In: An Evaluation and Interpretation of Reproductive Endpoints for Human Health Risk Assessment, G. Daston,. and C.A. Kimmel, (eds.), ILSI Press, Washington, DC.

[40] Regan, K. S., Cline, J. M., Creasy, D., Davis, B., Foley, G. L., Lanning, L., Latendresse, J. R., Makris, S., Morton, D., Rehm, S., and Stebbins, K. (2005). STP position paper: Ovarian follicle counting in the assessment of rodent reproductive toxicity. Toxicol. Pathol. 33, 409-412.

[41] Bolon, B., Bucci, T. J., Wargritton, A. R., Chen, J. J., Mattison, D. R., and Heindel, J. J. (1997). Differential follicle counts as a screen for chemically induced ovarian toxicity in mice: results from continuous breeding bioassays. Fundam. Appl. Toxicol. 39, 1-10.

[42] Hirshfield, A. N. (1987). Histological assessment of follicular development and its applicability to risk assessment. Reprod. Toxicol. 1, 71-79.

[43] Stütz, N., Bongiovanni, B., Rassetto, M., Ferri, A., Evangelista de Duffard, A. M., and Duffard, R. (2006). Detection of 2,4-dichlorophenoxyacetic acid in rat milk of dams exposed during lactation and milk analysis of their major components. Food Chem. Toxicol. 44, 8-16.

[44] Thomas, J. A. (1991). Toxic Responses of the Reproductive System. In: Casarett and Doull's Toxicology, M.O. Amdur, J. Doull, and C.D. Klaassen (eds.), Pergamon, New York.

[45] Zenick, H., Clegg, E.D. (1989). Assessment of Male Reproductive Toxicity: A Risk Assessment Approach. In: Principles and Methods of Toxicology, A.W. Hayes (ed.), Raven Press, New York.

[46] Palmer, A.K. (1981). In: Developmental Toxicology, Kimmel, C.A. and J. Buelke-Sam (eds.), Raven Press, New York.

[47] Palmer, A.K. (1978). In Handbook of Teratology, Vol. 4, J.G. Wilson and F.C. Fraser (eds.), Plenum Press, New York.

[48] Thigpen, J.E., K.D.R. Setchell, K.B. Ahlmark, J. Locklear, T. Spahr, G.F. Leviness, M.F. Goelz, J.K. Haseman, R.R. Newbold, and D.B. Forsythe (1999). Phytoestrogen content of purified open and closed formula laboratory animal diets. *Lab. Anim. Sci.* **49,** 530- 536.

[49] Gallavan et al., 1999, *Reprod. Toxicol.* 13: 383-390

[50] Bolin, B., R. Garman, K. Jensen, G. Krinke, B. Stuart, and an *ad Hoc* Working Group of the STP Scientific and Regulatory Policy Committee, "A 'Best Practices' Approach to Neuropathologic Assessment in Developmental Neurotoxicity Testing – for Today", *Toxicologic Pathology, 34*, 296-313, 2006.

[51] OECD Guidance Document n° 43 on Mammalian Reproductive Toxicity Testing and Assessment
http://www.olis.oecd.org/olis/2008doc.nsf/LinkTo/NT0000358A/$FILE/JT03249193.PDF

[52] Lanning et al "Recommended Approaches for the Evaluation of Testicular and Epididymal Toxicity." Toxicologic Pathology 30, 507-520, 2002

[53] Ladics GS. 2007. Use of SRBC antibody responses for immunotoxicity testing. Methods 41, 9-19

[54] Gore ER, Gower J, Kurali E, Sui JL, Bynum J, Ennulat D, Herzyk DJ. 2004. Primary antibody response to keyhole limpet hemocyanin in rat as a model for immunotoxicity evaluation. Toxicology 197, 23-35.

Appendix A

Measures and Observations Included in the Functional Observational Battery

Home Cage & Open Field	Manipulative	Physiologic
Posture	Ease of removal	Temperature	
Involuntary Clonic & Tonic	Ease of handling	Body weight
Palpebral Closure	Muscle Tone	Pupil response
Piloerection	Approach Response	Pupil size	
Salivation	Touch Response	
Lacrimation	Auditory Response	
Vocalizations	Tail Pinch Response	
Rearing	Righting Response	
Urination	Landing Foot Splay	
Defecation	Forelimb Grip Strength	
Gait Abnormalities	Hindlimb Grip Strength	
Arousal		
Stereotypy		
Bizarre Behavior		
Stains		
Respiratory Abnormalities		

Appendix B

Proposed Triggers in the Extended One-Generation Study for Producing a Second Generation (mating of the F_1 offspring to produce F_2 litters)

The F_1 extended study design includes evaluations of numerous sensitive structural, functional, and endocrine-mediated components. Thus, it is unlikely that any critical effect on development and reproduction would be missed. Using a science and risk based approach (as described in Cooper et al., 2006) to determine the need for an F_2 evaluation allows for tailored approach to testing, reduces the numbers of animals used (1200 animals are used to generate an F_2), and the resources needed to manage, review, and document the study. If deemed necessary, the production of an F_2 (i.e., breeding of the Cohort 1 F_1 animals) does represent a critical decision point integral to the study design. This decision will need to be made rapidly with a clear understanding of the data that supports it. When determining whether production of an F_2 is needed, one should consider how the additional information gained by breeding a second generation will be used in the safety evaluation or risk assessment. Typically, the F_2 generation allows for a replicate assessment of reproductive performance, litter size, offspring survival and development (including anogenital distance and nipple retention), and weanling necropsy endpoints (organ weights and histopathology).

Cooper *et al.* (2006) identified several triggers for the production of a second-generation as:

- an adverse effect on fertility or fecundity of the parental generation,
- indications of abnormal sexual development of the F_1 pups,
- deaths or evidence of toxicity to the F_1 pups preweaning.
- equivocal effects on F_1 parameters or unusual control data compared to historical background may also trigger a second generation

Further clarification and discussion of the potential triggers follows.

In-life results for the numerous endpoints examined in the P and F_1 offspring of the F_1 extended study should be considered when determining whether to conduct a second breeding. Also, existing knowledge from previously conducted reproductive and developmental toxicity studies and available mode of action information may contribute to the decision to mate the F_1 offspring.

If effects on neurotoxicity or immunotoxicity are the most sensitive outcomes, then breeding the F_1 offspring to produce an F_2 offers no advantage for risk assessment.

Additionally, assessments of gross pathology, organ weights and histopathology from the F_2 weanling necropsy would offer no advantage over the evaluation of the F_1 weanling animals because this cohort receives general toxicity evaluations, including histopathological and neuropathological evaluations as part of the proposed extended one-generation protocol. Moreover, the F_1 adults are exposed for a greater period than the F_2 weanlings (*in utero* through weaning and adulthood) and receive a far more comprehensive evaluation.

Reproductive toxicity is the scenario likely to trigger a breeding of the F_1 to produce the second generation. Table 1 lists the effects that could serve as potential triggers for the generation of the F_2 and whether the data would be available in time to make a decision.

Male Reproductive Endpoints: It is generally accepted that reproductive organ histopathology is the most sensitive endpoint for detecting minor changes in spermatogenesis in rats (Mangelsdorf *et al.,* 2003; Ulbrich and Palmer, 1995). Functional evaluations of fertility are less sensitive due to the excess sperm reserve in rodents. Mangelsdorf *et al.* (2003), in an assessment for the German Federal Institute on Occupational Safety and Health, reported that reproductive/accessory sex gland organ weights and sperm parameters (motility and counts) were more sensitive endpoints for detecting toxicant effects on reproduction than male fertility (number of implantations and pregnancies). These results are consistent with a limited analysis by Gray *et al.* (1989) who reported effects on sperm and gonadal toxicity occurred at lower doses than effects on fertility. These data consistently support the premise that alterations in sperm parameters will be more sensitive at detecting potential adverse effects than a second mating.

Given the greater sensitivity of histopathology and sperm evaluations to detect changes in male reproductive toxicity compared to the functional assessment by a second breeding, neither effects on reproductive histopathology nor effects on sperm parameters warrant breeding a second generation. These endpoints are assessed twice in the extended one-generation study, including an assessment of F_1 offspring that have been exposed *in utero*, during lactation and maturation. F_1 offspring data on reproductive organ weights, histopathology and andrology will not be available when a decision to conduct a second breeding is needed (Table 1); however, the greater sensitivity of these endpoints fulfills risk assessment needs better than a second breeding. Furthermore, neither mature testicular histopathology nor sperm assessment is assessed in F_2 animals, which are euthanized at weaning.

Female Reproductive Endpoints: With respect to the female endpoints (estrous cycle evaluation, reproductive organ weights and histopathology, and ovarian follicle counts), there are fewer data available for the comparison of endpoint sensitivity. Generally, it is recognized that alterations in ovarian follicle development in female rats may not affect fertility (Hirshfield, 1987). Female rats have robust reproductive performance and hormone production even in the presence of reproductive system alterations (*e.g.*, within 24 hours of removing an ovary, the remaining ovary can ovulate a full complement of 10-12 follicles; only 2-3 corpora lutea are needed to maintain pregnancy – Hirshfield, 1987). Thus, ovarian histopathology is believed to be a sensitive indicator of female reproductive toxicity (Regan *et al.,* 2005). Examination of adult ovarian histopathology is favored as decreases in primordial follicle numbers will be exacerbated in adults due to continuous recruitment of the remaining follicles (Regan *et al.,* 2005). Therefore, examination of adult P and F_1 offspring is preferred over an examination of prepubescent F_2 weanlings. While the Society of Toxicologic Pathology favors qualitative ovarian histopathology (conducted in the Ps) as an initial assessment of ovarian effects (Regan *et al.,* 2005), Bolon *et al.* (1997) has suggested that ovarian follicle counts (conducted in the F_1) also provide a more sensitive indicator of female reproductive toxicity than fertility, again favoring a histopathology endpoint over fertility.

In contrast to ovarian histopathology, estrous cyclicity was deemed "specific, but not sensitive" after an evaluation of the NTP continuous breeding study database (Chapin *et al.,* 1997). Estrous cycle length is variable across animals and subject to influence by stress (Matysek, 1989; Roozendaal *et al*., 1995). However, increased estrous cycle length has a slight association with decreased number of pups in the NTP database (Chapin *et al.,* 1997). Consequently, dose-related and biologically relevant changes in estrous cycle length that are not associated with overt maternal toxicity would warrant production of a second generation. Estrous cycle data from both the P and F_1 animals will be available to evaluate as a potential trigger for a second breeding.

With respect to reproductive organ weights, uterine weight is highly variable, depending on the stage of the estrous cycle at necropsy. (The stage of the estrous cycle at the time of necropsy is also not a predictive

endpoint, being limited to a single timepoint.) A limited assessment by Gray *et al.* (1988) indicated that ovarian and pituitary weights had similar sensitivity to fertility and litter size in detecting one estrogenic chemical. In this assessment, puberty onset and percent of animals with normal cycles were among the most sensitive endpoints, although reproductive organ histopathology was not conducted. Again, one should consider data availability and endpoint sensitivity in the context of the extended one-generation toxicity study when determining whether a second breeding is needed (Table 1). Data on P fertility, litter parameters, F_1 developmental landmarks, and P reproductive organ weights and histopathology would be available for decision making. Effects on reproductive organ weights and histopathology do not warrant a second mating as these endpoints are either more or as sensitive to toxicant alterations than fertility and because additional information on these endpoints would not be obtained in F_2 pups. A second assessment of reproductive organ weights and histopathology will be available from the F_1 offspring, including ovarian follicle counts. These data should fulfill risk assessment needs.

Other Endpoints: In contrast, effects on F_1 litter size in the absence of P reproductive organ histopathology changes or effects on pup survival in the absence of overt severe maternal toxicity or on pup developmental landmarks (discussed below) would require a second breeding.

While puberty onset is only examined in F_1 offspring, the advantage of the extended one-generation study is that more animals from each litter are assessed for puberty onset (3 sex/litter versus 1 sex/litter). While these data are analyzed by litter, compiling data from more pups per litter will ensure that more accurate values are used to calculate mean age at puberty onset and variability will be reduced. If dose-related alterations in puberty onset occur that are not secondary to body weight effects, breeding of a second generation is warranted.

There are some endpoints that, without a second breeding, are only examined in the F_1 offspring (Table 1). These include reproductive performance, litter size, offspring survival, offspring development (including anogenital distance and nipple retention) and endpoints assessed at weanling necropsy. Treatment-related alterations in these endpoints, including dose-related effects on P reproductive performance in the absence of treatment-related reproductive organ histopathology, justifies the production of a second generation, particularly if these findings occur in the absence of overt and severe maternal toxicity. This approach is consistent with the triggers outlined by Cooper *et al.* (2006) who identified: 1) an adverse effect on fertility or fecundity of the parental (P) generation; 2) indication of abnormal sexual development of the F_1 pups; and 3) deaths or evidence of toxicity to the F_1 pups preweaning. Triggers for production of the second generation should be based on toxicological significance and dose-related responses. As with other toxicological studies, weight of evidence will be applied when interpreting the results of the extended one-generation study. The weight of evidence concept becomes especially important given the number of endpoints examined and the opportunity for Type I error. The laboratory's historical control data (HCD) can assist in the interpretation of data on reproductive toxicity endpoints.

Decreased F_1 birth weight has been raised by some as a potential trigger to generate an F_2 evaluation. However, pup body weight changes in the absence of any other effect is too nonspecific for triggering an F_1 mating. Laws *et al*. (in press) has examined the effect of decreases in pup body weight (2-20% reductions) in rat pubertal assays and found that decreases less than 10% were without male or female reproductive or thyroid effects. With reductions greater than 10%, other changes are found in several parameters that are evaluated in this F_1 extended protocol and would serve as triggers (*e.g.,* delay in puberty) for the generation of an F_2. Although there may be body weight reductions that accompany mechanisms of action that could lead to potential reproductive/developmental concerns (*e.g.,* estrogenic acting compounds can affect appetite and thus potentially impact body weight), there are more specific and sensitive reproductive effects that would be identified (*e.g.,* advanced vaginal opening in the F_1, ovarian

cycling on the P and F_1 generation). The triggers identified in Table 1 are more specific in further characterizing the impact of chemicals on fertility and reproductive development.

Lastly, when potential human exposures are considered adequately characterized, Margin of Exposure (MOE) considerations could be factored into the decision to require a second generation breeding. For example, if toxicity triggers are limited to the high dose level alone (with no apparent dose-related trend), margins of exposure of this dose relative to either estimated human exposures or those directly measured through human biomonitoring studies could help guide the triggering decision.

Table B1: Availability of Data When Deciding About a Second Breeding

Endpoint	Available to make an F2 decision	Comments
P Estrous Cycle Evaluation	Yes	Trigger[1]
P Fertility	Yes	Trigger[2]
F_1 Litter parameters	Yes	Trigger[3]
F_1 Developmental Landmarks (AGD, nipple retention, puberty onset)	Yes	Trigger[4]
P Reproductive Organ Weights	Yes	
P Reproductive Organ Histopathology	Yes	Endpoints are more sensitive than fertility; second breeding not needed
P Andrology (Sperm Parameters)	Yes	
P Qualitative Ovarian Assessment	Yes	
F_1 Estrous Cycle Evaluation	Yes	Trigger[1]
F_1 Reproductive Organ Weights	No	
F_1 Reproductive Organ Histopathology	No	Not applicable triggers, but suitable for hazard characterization
F_1 Andrology (Sperm Parameters)	No	
F_1 Qualitative Ovarian Assessment	No	

[1] If biologically relevant, dose-related changes in estrous cycle length without overt toxicity in the dams

[2] In the absence of corresponding, treatment-related reproductive organ histopathology

[3] If significant, treatment-related decreases in litter size/pup survival are seen in the absence of severe maternal toxicity or lethality

[4] Dose-related effects; in the absence of body weight-mediated changes in these parameters

~~**Proposed Triggers in the Extended One-Generation Study for Including a Cohort 3 (Immunotoxicity)**~~

~~A reduction in serum T4 levels in dams at weaning of their litters after treatment with PTU from day 10 of pregnancy to day 21 of lactation was associated with a reduced immune function in offspring mice (Watanabe *et al*., 2008). Any indication that the test substance exerts an effect on the thyroid could be considered a trigger for a more detailed evaluation of developmental immunotoxicity.~~

REFERENCES

Bolon, B., Bucci, T. J., Wargritton, A. R., Chen, J. J., Mattison, D. R., and Heindel, J. J. (1997). Differential follicle counts as a screen for chemically induced ovarian toxicity in mice: results from continuous breeding bioassays. *Fundam. Appl. Toxicol*. 39, 1-10.

Cooper, R. L., Lamb, J. C., Barlow, S. M., Bentley, K., Brady, A. M., Doerr, N., Eisenbrandt, D. L., Fenner-Crisp, P. A., Hines, R. N., Irvine, L. F. H., Kimmel, C. A., Koeter, H., Li, A. A., Makris, S. L., Sheets, L. P., Speijers, G. J. A., and Whitby, K. E. (2006). A tiered approach to life stages testing for agricultural chemical safety assessment. *Crit. Rev. Toxicol*. 36, 69-98.

Gray, L. E., Jr., Ostby, J., Sigmon, R., Ferrell, J., Rehnberg, G., Linder, R., Cooper, R., Goldman, J., and Laskey, J. (1988). The development of a protocol to assess reproductive effects of toxicants in the rat. *Reprod. Toxicol*. 2, 281-287.

Gray, L., Ostby, J., Ferrell, J., Sigmon, R., Cooper, R., Linder, R., Rehnberg, G., Goldman, J., and Laskey, J. (1989). Correlation of sperm and endocrine measures with reproductive success in rodents. In: Progress in Clinical and Biological Research, vol. 302 (E.J. Burger, ed.) Liss, New York, pp. 193-209.

Hirshfield, A. N. (1987). Histological assessment of follicular development and its applicability to risk assessment. *Reprod. Toxicol.* 1, 71-79.

Mangelsdorf, I., Buschmann, J., and Orthen, B. (2003). Some aspects relating to the evaluation of the effects of chemicals on male fertility. *Regul. Toxicol. Pharmacol.* 37, 356-369.

Matysek, M. (1989). Studies on the effect of stress on the estrus cycle in rats. Ann. Univ. Mariae Curie Sklodowska 44, 143-149.

Regan, K. S., Cline, J. M., Creasy, D., Davis, B., Foley, G. L., Lanning, L., Latendresse, J. R., Makris, S., Morton, D., Rehm, S., and Stebbins, K. (2005). STP position paper: Ovarian follicle counting in the assessment of rodent reproductive toxicity. *Toxicol. Pathol*. 33, 409-412.

Roozendaal, M. M., Swarts, H. J., Wiegant, V. M., and Mattheij, J. A. (1995). Effect of restraint stress on the preovulatory luteinizing hormone profile and ovulation in the rat. *Eur. J. Endocrinol*. 133, 347-353.

Ulbrich, B., and Palmer, A. K. (1995). Detection of effects on male reproduction – a literature survey. *J. Am. Coll. Toxicol*. 14 (4), 293-327.

Watanabe, W., Shimizu, T., Hino, A., and Kurokawa, M. (2008). A new assy system for evaluation of developmental immunotoxicity of chemical compounds using respiratory syncytial virus infection to offspring mice. *Environ. Toxicol. Pharmacol.* 25, 69-74.

Appendix C

Endpoints to be evaluated as part of the Extended One-Generation Reproductive Toxicity Guideline

Item	Proposed Addition(s)	Comments and Considerations
1	Estrous Cyclicity	
	P and females	• Two weeks prior to start of exposures (only put cycling P females on study) • During 2-week prebreed exposure period for P (and F_1?) (to see if agent affects EC) • During P/F_1 mating until female is sperm or copulation plug positive. Each day during the mating period, the females should be examined for the presence of sperm or vaginal plugs. The number of day s until the plug is observed should be analyzed as an indirect indicator of sexual behavior.
	F_1 females	• For 1-2 weeks after vaginal patency (to identify time at which estrous cycling is initiated in F_1) • During the last 3 weeks of the postwean exposure period in F_1 (to see if agent affects EC) • If the F_1 animals are mated to generate the F_2 offspring, during cohabitation until evidence of insemination • If the F_1 animals are mated, after the weaning (and necropsy) of the F_2 offspring for ~1 week in F_1 (to see if EC is reinstated)
	P and F_1 females	Stage of estrus at demise for P and F_1 (necropsy; compromise from necropsying all females on same stage of estrus)
2	**At necropsy** of P and F_1 dams in addition to current list: examine grossly, weigh, and retain ovaries, uterus, and cervix and vagina in fixative. Present weights as absolute and relative to body (and brain?) weight	If appropriate, count implantation sites to calculate postimplantation loss
	Endpoints in F_1 females if masculinization is indicated	Body weight, any unusual malformation or anomalies, count nipples and areolas (observed blind to treatment), record position of areolas and nipples, uterine abnormalities, including bi-or uni- lateral agenesis of oviducts, uterine horns, infections, hydrometrocolpous etc.
3	**Histopathology** initially on high dose and control retained tissues from P and F_1 females (randomly selected, 10/group)	• If an apparent treatment-related effect is observed, perform histopathology on the organ of interest in the lower dose groups (ALL) and perform histopathology on all

		remaining organs of interest in high dose and control females • Step section ovaries and count the stages of development from primordial follicles to Graafian follicles (10/group, high dose and control; see above for progression)
4	**Acquisition of puberty** (vaginal patency) for F_1 females (absolute age and covaried by body weight at acquisition and by body weight on a fixed day; *e.g.,* pnd 30)	• Include if vaginal thread • Begin observations on pnd 25 for CD (SD) rats unless the test material is known to have estrogenic activity (then start examinations earlier)
5	**Anogenital distance between pnd 0 and pnd4**, and if an effect is noted, also on pnd 21 (at weaning), and at adult necropsy (absolute and covaried by body weight at measurement)	Probably most sensitive male measurement for anti-androgens
6	In addition to standard endpoints (body weight, feed consumption, clinical observations, including mortality/morbidity checks): (1) date of pairing, (2) date of insemination, and (3) date of delivery	• To calculate precoital interval (pairing to insemination) • To calculate gestational length (insemination to delivery)
7	**Data for F_1 and F_2 progeny for pnd 0-21 (wean)**:	
	• Number of pups (total/live/dead), individual sex, gross examination, body weight, and anogenital distance (uniquely identify by paw tattoo); on PND 0	• On PND 0 (or however date of birth is designated), 4, 7, 14, and 21
	• Number of pups (total/live/dead), individual sex, gross examination, body weight; on PND 4, 7, 14, and 21	
	• On PND 4: standardize pups to 10 (sex ratio 5:5, 6:4, 4:6) for CD (SD) or to 8 (4:4, 5:3, 3:5) for strains with smaller litter size	• To remove the confounder of litter size on survival, growth, body weights, acquisition of developmental landmarks, etc.
	• Developmental landmarks	• *e.g.*, pinna detachment, surface righting reflex, pilation, eye opening, acquisition of auditory (acoustic) startle (opening of auditory canal), incisor eruption, testis descent (usually pnd 15-21)
	• PND 21 females: For F_1 and F_2 weanling culls (if any), histopathological examination of treatment-related abnormalities noted at macroscopic examination should be considered, if such evaluation were deemed appropriate it would contribute to the interpretation of the study data	• See 2 and 3 above for details
8	**Weaning** (pnd 21) F_1 offspring, anogenital distance, body weight	• All continue on study (based on discussions at EPA on 11/8-9/06 see

		Attachment 2)
	F_2 offspring: necropsy anogenital distance and body weight at weaning, necropsy, organ weights and retain in appropriate fixative Organ weights reported as absolute and relative to body (and brain?) weight Histopathology of retained tissues initially in 10/sex/group and only high dose and control groups (see #3)	• Systemic organs (both sexes): brain, CNS, PNS (dorsal root ganglia), liver, kidneys, spleen, adrenal glands, pituitary (weigh postfixation), thyroid (weight postfixation), thymus, GI tract, trachea, lungs, urinary bladder, bone marrow, eye plus optic nerve; identified target organs • Plus male: testes, epididymides, seminal vesicles with coagulating glands, prostate • Plus female: ovaries, uterus, cervix and vagina

Proposed Additions to Male Endpoints

Item	Proposed Addition(s)	Comments and Considerations
1		
	P males: Dosing In 10 week dosed PO males, do standard necropsy when at time F1 pups are weaned with careful attention to male reproductive tract and andrology, thyroid hormones and histology of the reproductive tract.	• 4-10 weeks prior to start of exposures (see text for details)
	$F_1$and F_2 males, <u>Anogenital distance, all animals</u> <u>Areola/nipples</u>	What, where and how many, in both males and females on PND 13
	<u>Acquisition of puberty</u> (preputial separation) for all F_1 males (absolute age and covaried by body weight at acquisition and by body weight on a fixed day; *e.g.*, PND 42)	• Note and record any threads • Begin observations on PND 42 for CD (SD) rats <u>unless</u> the test material is known to have androgenic activity (then start examinations earlier)
2	**<u>At necropsy</u>** of P and F_1 males, in addition to listed endpoint organs, histopathological examination of treatment-related abnormalities noted at macroscopic examination should be considered to assist in data interpretation. TSH, T4, thyroid weight, thyroid histology, all at necropsy	Record testis weight (individually one for sperm numbers and one for histology). Weigh corpus plus caput epididymides and cauda epididymides (one for sperm one for histology). Weigh seminal vesicle, plus coagulating glands with fluid as a unit. Weigh kidneys, adrenals, liver, levator ani plus bulbocavernosus, Cowper's gland (as a pair) and glans penis

		Record testis location (descended, undescended, attached, floating). Note malformation, agenensis or inappropriate presence of any of the sex organs (*e.g.*, is prepuce partially or entirely detached from glans penis, prostate agenesis, presence of uterus in male). Count nipples and areolas and record position of each. Record prostate weight by lobe (ventral and dorsolateral) Note if prepuce is partially or entirely detached from glans penis, note presence of persistent thread Note if inguinal regions are soiled with urine. Note if gubernacular cords or cranial suspensory ligaments are present or absent Examine prostate, seminal vesicles and coagulating glands are small or absent, infected or one side larger than the other. Note if kidneys display hydronephrosis, calcium deposits and the presence of hydroureter, bladder stones or blood.
3	**Histopathology** initially on high dose and control retained tissues from P and F_1 males (randomly selected, 10/group)	• If an apparent treatment-related effect is observed, perform histopathology on the organ of interest in the lower dose groups (ALL) and perform histopathology on all remaining organs of interest in high dose and control males
5	**Anogenital distance on between pnd 0 and pnd4** , and if an effect is noted, also on PND 21 (at weaning), and at adult necropsy (absolute and covaried by body weight at measurement)	Probably most sensitive male measurement for anti-androgens
6	In addition to standard endpoints (body weight, feed consumption, clinical observations, including mortality/morbidity checks): (1) date of pairing, (2) date of insemination*, and (3) date of delivery *	• To calculate precoital interval (pairing to insemination) • To calculate gestational length (insemination to delivery)

Appendix D

Relevant data that can be obtained from repeat-dose studies

General information

- Selection of dosages
- Food and water consumption
- Toxic response data by sex and dose level
- Nature, severity and duration of clinical observations (whether reversible or not)
- Sensory activity, grip strength and motor activity assesments
- Data on ADME, if available and other toxicokinetic information
- Tissue weights

Haematology data with relevant base-line values

- Haematocrit
- Haemoglobin concentration
- Erythrocyte count
- Total and differential leucocyte count
- Platelet count
- Blood clotting time/potential

Clinical biochemistry

- Sodium, potassium, glucose, total cholesterol, urea, creatinine, total protein, albumin, enzymes indicative of hepatocellular effects (*e.g.* alanine aminotransferase, aspartate aminotransferase, alkaline phosphatise, gamma glutamyl transpeptidase, sorbitol dehydrogenase)
- Other parameters, as far as available (*e.g.* specific hormones, methaemoglobin, cholinesterase)

Reproductive system

- Tissue weight and tissue histopathology (gonads, accessory sex organs (*e.g.* uterus, prostate)
- Spermatogenesis (testicular histopathology)
- Sperm production
- Sperm integrity/function
- Oestrus cycle
- Follicle counts/oocyte maturation

- Ovarian integrity (histopathology)

Immune system

- Tissue weight / histopathology /gross necropsy of *e.g.* thymus, spleen, large intestines with Peyer's patches, lymph nodes, bone marrow

Central and peripheral nervous system

- Tissue weight / histopathology / gross necropsy of brain (and representative regions thereof), spinal cord, peripheral nerves