Draft version 10, July 2009

DRAFT Extended One-Generation Reproductive Toxicity TEST GUIDELINE

OECD/OCDE XXX

OECD GUIDELINE FOR THE TESTING OF CHEMICALS

INTRODUCTION

1. This Test Guideline (TG) has been designed to address the common goals of reduced/refined animal testing for toxicity to reproduction and offspring development, while preserving/enhancing 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 one generation reproductive study as published in Cooper et al., 2006 [1]. Several improvements and clarifications have been made to the study design to provide flexibility and to stress the importance of starting with existing knowledge, while using in-life observations to guide and tailor the testing. This guideline provides a detailed description of the operational conduct of an Extended One-Generation Reproductive Toxicity Study and lists a set of endpoints that may 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; this cohort may be extended to include an F_2 generation, depending on available data including observations during the in-life part of the study.
- Cohort 2: evaluates the potential impact of chemical exposure on the developing nervous system.
- Cohort 3: evaluates the potential impact of chemical exposure on the developing immune system.

INITIAL CONSIDERATIONS AND OBJECTIVES

2. The main objective of the Extended One-Generation Reproductive Toxicity Study is to evaluate specific life stages not covered by other types of toxicity studies[1] and test for effects that may occur as a result of pre- and postnatal chemical exposure. For reproductive endpoints, it is envisaged that, as a first step, repeat-dose studies (including screening reproductive toxicity studies, *e.g.* TG 422, or short term endocrine disrupter screening assays, *e.g.* pubertal assay) are used to detect effects on reproductive organs for males and females. This would include

[1] **NOTE**: Life stages not covered by other types of toxicity testing may include mating, pregnancy, parturition, lactation, offspring development prenatal to adult

spermatogenesis (testicular histopathology) for males and oestrous cycles, follicle counts/oocyte maturation and ovarian integrity (histopathology) for females. The Extended One-Generation Reproductive 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, and the F_1 generation throughout the life stages[2].

3. The TG is designed to provide an evaluation of the pre- and postnatal effects of chemicals on development as well as a thorough evaluation of systemic toxicity 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. 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. Specifically, but not exclusively, the following parameters are considered: gonadal function, the oestrous cycle, epididymal sperm maturation, mating behaviour, conception, pregnancy, parturition, and lactation[3]. Furthermore, the information obtained from the neurotoxicity and immunotoxicity assessments may better characterize potential effects in those systems. The data derived from these tests should allow the determination of No-Observed Adverse Effect Levels (NOAELs), Lowest Observed Adverse Effect Levels (LOAELs) and/or benchmark doses[4] for the various endpoints and/or serve as a guide for subsequent testing.

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; but for a minimum of two weeks) and a two-week mating period. P males are further treated at least until weaning of the F_1 (i.e. a minimum of 10 weeks of treatment). They may be treated for longer if there is a need to clarify effects on reproduction. Treatment of the P females is continued during pregnancy and lactation until termination after the weaning of their litters (i.e. 8-10 weeks of treatment). The F_1 offspring receive further treatment with the test substance from weaning to adulthood until postnatal day (PND) 70 or 90, depending on their 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.

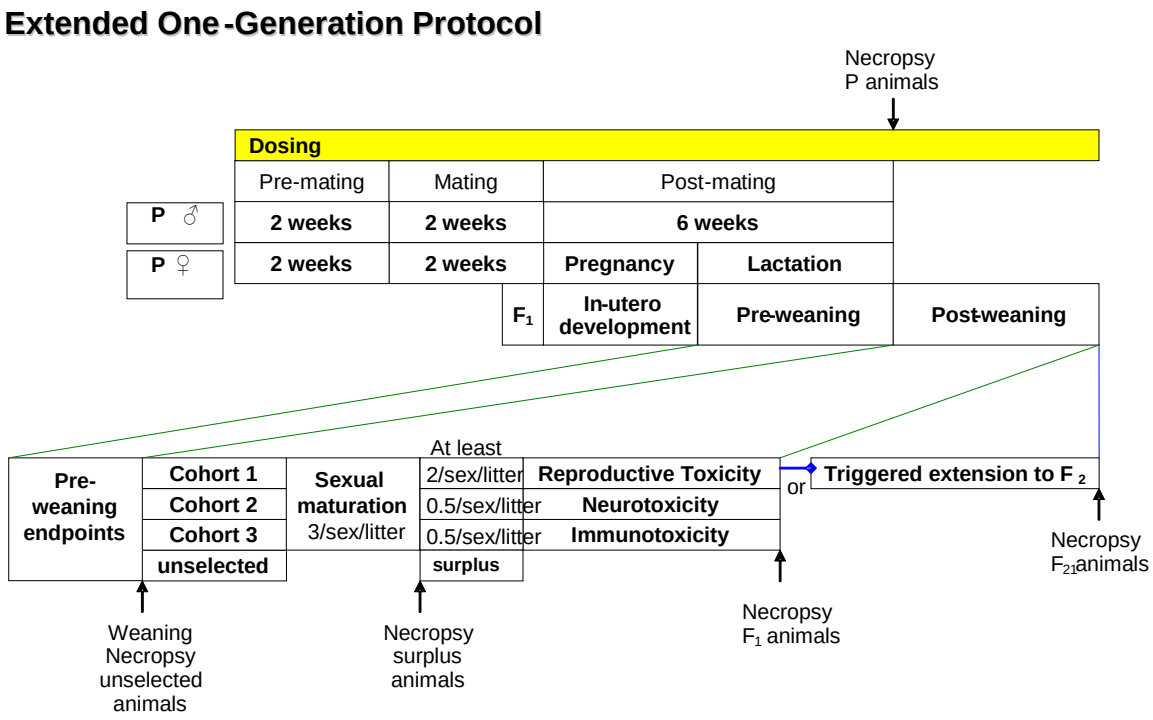
[2] **NOTE: Endpoints covered**
This includes mating performance and success (e.g.sperm maturation, behavioral parameters, secondary sex organ function), pregnancy (e.g. fertilisation, pre-implantation embryo transport to the uterus, priming of the uterine epithelium for implantation, placentation, hormonal 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.

[3] **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.

[4] **NOTE: Benchmark dose**
When the nature of the data do not allow clear determination of a NOAEL, the 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 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.

5. Clinical observations 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 (cohorts 1-3, see paragraphs 30 to 32 and Figures 1 & 2) for further investigations, including sexual maturation, reproductive organ integrity and function, neurological and behavioural endpoints, and immune functions.

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



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 should be carefully considered in 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 another species is 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, of similar weight at initiation of dosing, similar age (approximately 90 days) at mating, and representative of the species and strain under study. It is recommended that delivery of P females and P males to the test facility occurs at 8-10 weeks of age. Animals should be acclimated for at least 5 days after arrival. 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 (i.e. +/- 20% of the mean).

Housing and feeding conditions

8. The temperature in the experimental animal room should be 22 °C (+/- 3°). Relative humidity should be between 30-70 % with an ideal range of 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 phytoestrogen content. Standardized, open-formula diets in which estrogenic substances have been reduced are recommended[5] [2].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 should 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 suitable cages. After evidence of copulation, females that are presumed to be pregnant are housed separately in parturition or maternity cages where they are provided with appropriate and defined nesting materials when parturition is near (gestation days 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 ensure 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. *deleted*

12. Each P animal is assigned a unique identification number before dosing starts. If laboratory historical data suggest that a significant proportion of females may not show regular (4 or 5-day) oestrous cycles, then an assessment of oestrous cycles before start of treatment is advised. Alternatively, the group size may be increased to ensure that at least 20 of the control females would have regular (4 or 5-day) oestrous cycles at start of treatment. If P females are monitored for oestrous cycles before treatment is started, the females should receive their identification before this procedure is initiated. All F_1 offspring are uniquely identified when neonates are first

[5] **NOTE**: Phytoestrogen content should be preferably reduced to levels that will not alter study results.

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 potential modifications of the dosing schedule. Therefore, all the relevant available information on the test substance, i.e. physico-chemical, TK (including species-specific metabolism), toxicodynamic properties, structure-activity relationships (SARs), *in vitro* metabolic processes, results of previous toxicity studies and relevant information on structural analogues should be taken into consideration in planning the Extended One-Generation Reproductive Toxicity Study. Preliminary information on absorption, disposition, metabolism and elimination (ADME) and bioaccumulation may be derived from chemical structure, physico-chemical data, extent of plasma protein binding or TK studies, while results from toxicity studies give additional information, *e.g.* on NOAEL, metabolism or induction of metabolism.

Consideration of toxicokinetic data

13 (*bis*). Although not required, TK data from previously conducted dose range-finding or other studies are extremely useful in the planning of the study design, selection of dose levels and interpretation of results. Of particular utility are data which: 1) verify exposure of developing fetuses and pups to the test compound (or relevant metabolites), 2) provide an estimate of internal dosimetry, and 3) evaluate for potential dose-dependent saturation of kinetic processes. Additional TK data, such as metabolite profiles, concentration-time courses, etc. should also be considered, if they are available. Supplemental TK data may also be collected during the main study, provided that it does not interfere with the collection and interpretation of the main study endpoints.

As a general guide, the following represents a modest TK data set which would be useful in planning the Extended One-Generation Reproductive Toxicity Study:

- Late pregnancy (*e.g.*, GD20) - maternal blood and foetal blood
- Mid-lactation (PND 10) - maternal blood, pup blood and/or milk
- Early post-weaning (*e.g.*, PND28) - weanling blood samples

Flexibility should be employed in determining the specific analytes (*e.g.* parent compound and/or metabolites) and sampling scheme. For example, the number and timing of sample collection on a given sampling day will be dependent upon route of exposure and prior knowledge of TK properties in non-pregnant animals. For dietary studies, sampling at a single consistent time on each of these days is sufficient, whereas gavage dosing may warrant additional sampling times to obtain a better estimate of the range of internal doses. However, it is not necessary to generate a full concentration time-course on any of the sampling days. If necessary, blood can be pooled by sex within litters for fetal and neonatal analyses.

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 should 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.

Dose Selection

16. Normally, the study should include at least three dose levels and a concurrent control. When selecting appropriate dose levels, the investigator should consider all available information, including the dosing information from previous studies, TK data from pregnant or non-pregnant animals, the extent of lactational transfer, and estimates of human exposure. If TK data are available which indicate dose-dependent saturation of TK processes, care should be taken to avoid high dose levels which clearly exhibit saturation, provided of course, that human exposures are expected to be well below the point of saturation. In such cases, the highest dose level should be at or just slightly above the inflection point for transition to nonlinear TK behaviour.

In the absence of relevant TK data, the dose levels should be based on toxic effects, unless limited by the physical/chemical nature of the test substance. If dose levels are based on toxicity, the highest dose should be chosen with the aim to induce some systemic toxicity, but not death or severe suffering of the animals.

A descending sequence of dose levels should be selected in order to demonstrate any dose-related effect and to establish NOAELs or doses near the limit of detection that would allow for derivation of 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 doses.

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. If there is no evidence of toxicity at a dose of at least 1000 mg/kg body weight/day in repeat-dose studies, or if toxicity would not be expected based upon data from structurally- and/or metabolically-related compounds, indicating similarity in the *in vivo*/*in vitro* metabolic properties, a study using several dose levels may not be necessary. In such cases, the Extended One-Generation Reproductive Toxicity Study could be conducted using a control group and a single dose 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 only when human exposure does not indicate the need for a higher dose level.

PROCEDURES

Exposure of offspring

19, 20, 21 – *deleted. A merged paragraph (13 bis) moved to section just below Paragraph 13*.

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 during the lactation period should be considered based on available TK 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 oestrous cycles, male and female reproductive tract histopathology and testicular/epididymal sperm analysis may be available from previous repeat-dose toxicity studies of adequate duration. The duration of the pre-mating treatment in the Extended One-Generation Reproductive Toxicity Study therefore is aimed at the detection of effects on 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 oestrous 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 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[6].

24. Pre-mating exposure scenarios for males could be adapted, 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[7].

---

[6] **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.

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

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 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, depending on cohort assignment (see Figure 2).

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 should be specified.

27. When the test substance is administered by gavage, 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 F1 animals during the 2 weeks following weaning. 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, unrelated male from the same dose group (1:1 pairing) until evidence of copulation is observed or 2 weeks have elapsed. If there are insufficient males, for example due to male death before pairing, then male(s) which have already mated may be paired (1:1) with a second female(s) such that all females are paired. Day 0 of pregnancy is defined as the day on which mating evidence is confirmed (a vaginal plug or sperm are found). Animals should be separated as soon as possible after evidence of copulation is observed. If mating has not occurred after 2 weeks, the animals should be separated without further opportunity for mating. Mating pairs should be clearly identified in the data.

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.

Litter size

29. Standardization of litter size to 10 is recommended in order to remove any possible effect on litter size on parameters such as survival, growth, body weight, and acquisition of developmental landmarks. 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 and consideration given to measuring serum thyroid hormone (T4) concentrations. If necessary, neonatal (PND 4) blood can be pooled by litters for biochemical/thyroid hormone analyses.

Selection of pups for post-weaning studies (see Figure 2)

30. If standardization is not performed on PND4, at weaning (around PND 21) at least 3 male and 3 female pups (if possible) from each litter are selected for further examinations and maintained until sexual maturation (unless earlier testing is required). 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.

On PND21, the selected F_1 pups are randomly 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

<u>Cohort 1: Reproductive Toxicity Testing (PND 90)</u>

Cohort 1 pups (2 pups/sex/dose per litter, 40/sex/dose) are examined for sex-specific maturational endpoints (preputial gland separation, vaginal opening).

Cohort 1A: One animal/sex/dose/litter (20/sex/dose) have all organs weighed and examined for histopathology (as outlined in paragraph 64).

Cohort 1B: One animal/sex/dose/litter (20/sex/dose) have reproductive organs weighed and corresponding tissues processed to the block stage, as follows:

- Vagina
- Uterus with cervix
- Ovaries, testes (at least one)
- Epididymides
- Seminal vesicles
- Prostate and coagulating glands
- Identified target organs

Histopathology in cohort 1B would be conducted in the following situations:

- in cases of suspected reproductive or endocrine toxicants, based on chemical structure similarities or evidence from *in vivo* or *in vitro* studies or data generated at earlier time points, reproductive organs and endocrine tissues should be examined for histopathology
- if results from cohort 1A are equivocal

Cohort 2: Developmental Neurotoxicity Testing (PND21 [optional] and young adults)
Ten animals/sex/dose are terminated on PND 21 (optional, from the animals unselected at weaning) and/or 10 animals/sex/dose are terminated after puberty and after behavioural testing, and have organs weighed with full histopathology for purposes of neurotoxicity assessment. Cohort 2 animals are also examined for sex-specific maturational endpoints (preputial gland separation, vaginal opening).

Cohort 3: Developmental Immunotoxicity Testing (PND70)
10 animals/sex/dose are immunized at PND 42-70, and terminated 4 or 5 days after immunization. All animals have their organs weighed. Cohort 3 animals are also examined for sex-specific maturational endpoints (preputial gland separation, vaginal opening).

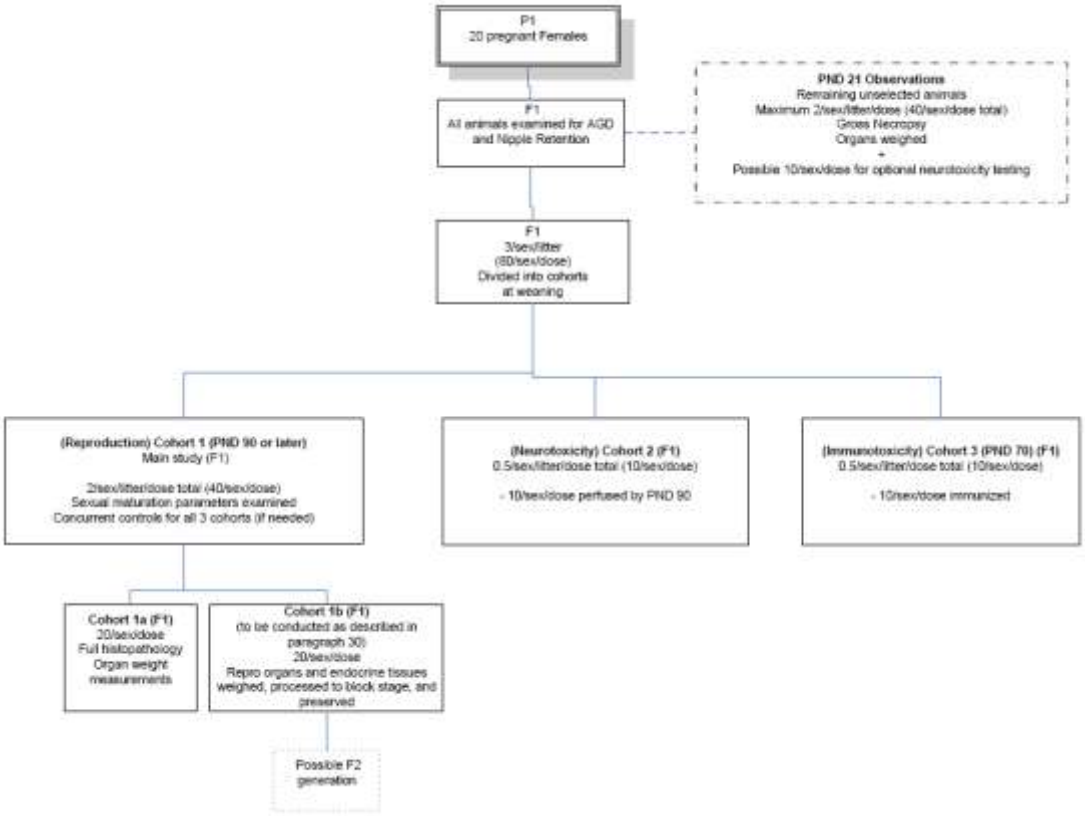
Animals not selected for cohorts:
The pups not selected for cohorts, including runts, are terminated after weaning, unless the results indicate the need for further in-life investigations. Terminated pups are subjected to gross necropsy on reproductive organs, as described in paragraph 65. Specified organs are weighed and preserved for possible histopathological examinations. Serum thyroid hormones (T4 and TSH) are measured. If needed, these animals will also be used for developmental neurotoxicity assessment at PND 21.

31. *relocated*

32. Should there be an insufficient number of pups in a litter to serve all cohorts, the cohort 1 takes precedence as it can be extended to produce an F_2 generation. Additional pups 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. These pups may be used for examinations at different timepoints, for the evaluation of supplementary endpoints, or to increase sample size in the cohort.

33. *deleted*

Figure 2: Overview of cohorts



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 once a day. In the case of gavage dosing, the timing of clinical observations should be prior to and post dosing (for possible signs of toxicity associated with peak plasma concentration). 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 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 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. 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) 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 PND 30 and PND 50). 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) are 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 to respective cohort.

Clinical biochemistry / Haematology

39. *deleted*

40. When monitoring for systemic effects, fasted blood samples from a defined site are taken from ten randomly selected P males and females per dose group at termination, stored under appropriate conditions and subjected to partial or full scale haematology[8], clinical biochemistry[9], assay of T4 and TSH or other examinations suggested by the known effect profile of the test substance[10]. In addition, blood from all animals may be taken and stored for possible analysis at a later time 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

[8] **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.

[9] **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.

[10] **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).

collected a few days before the animals are mated for the second time. Urinalysis may be performed prior to termination.[11]

41. For the investigation of pre- and postnatally induced major toxic (functional) effects on non-reproductive tissues in F1 animals, 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 to the following at termination:

- standard clinical biochemistry, including the assessment of serum levels for thyroid hormones (T4 and TSH),
- haematology (total and differential leukocyte plus erythrocyte counts),
- weighing of the adrenal glands and of the spleen,
- splenic lymphocyte subpopulation analysis (CD4+ and CD8+ T lymphocytes, B lymphocytes, and NK cells) using one half of the spleen,
- histopathology on the remaining half of the spleen,
- urinalysis assessments.

At termination, spleens and adrenals from cohort 1 animals will also be weighed, and one half of the spleen will be used for analysis of CD4+ and CD8+ T lymphocytes, B lymphocytes and natural killer cells.

Oestrous cycles

42. Preliminary information of test substance-related effects on the oestrous 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 Reproductive Toxicity Study. Normally the assessment of oestrous cyclicity (by vaginal cytology) will start at the beginning of the treatment period and continue until confirmation of mating or the end of the 2-weeks mating period. If females have been screened for normal oestrous cycles before treatment, then it is useful to continue smearing as treatment starts, but if there is concern about non-specific effects at the start of treatment (such as an initial marked reduction in food consumption) then animals may be allowed to adapt to treatment for up to two weeks before the start of the 2-weeks smearing period leading into pairing. If the female treatment period is extended in this way (i.e. to a 4-weeks pre-mating treatment) then consideration should be made to purchasing animals younger and to extending the period of male treatment before pairing. When obtaining vaginal/cervical cells, care should be taken to avoid disturbance of mucosa and subsequently, the induction of pseudopregnancy [11, 12].

43. *deleted*

44. Vaginal smears should be examined daily for all F_1 females in cohort 1A, after the onset of vaginal patency, until the first oestrus, in order to determine the time interval between these two events. Oestrous cycles for all F1 females in cohort 1A should also be monitored for a period of two weeks commencing around PND 75. In addition, should mating of the F1 generation be

[11] **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).

necessary, the vaginal cytology in cohort 1B will be followed from the time of pairing until mating evidence is detected.

45. For P and F_1 females, the oestrus stage is determined at termination to allow correlation with histopathology in reproductive organs and hormone measurements, if applicable.

Mating and pregnancy

46. In addition to the standard endpoints (body weight, food consumption, clinical observations including mortality/morbidity checks), the dates of pairing, the date of insemination and the date of parturition are recorded and the precoital interval (pairing to insemination) and the duration of pregnancy (insemination to parturition) are calculated. The P females should be examined carefully at the time of expected parturition for any signs of dystocia. Any abnormalities in nesting behaviour 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 parturition (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, state of activity and reaction to handling. 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 when the offspring are weighed, or more often if case specific findings have been made at birth. Signs noted could include, but may 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 (AGD) of each pup should be measured between PND 0 and PND 4. 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 [13]. The presence of nipples/areolae in male pups should be checked on PND 12 or 13.

50. *# to be revisited when the DNT endpoints are agreed #* Physical or behavioural abnormalities observed in the dams or offspring should be recorded. Evaluation of 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) should be conducted as they may give supplementary information and serve as markers for endocrine effects and nutritional deficiencies.

51. To allow for early detection of vaginal patency, at least three F_1 females per litter are evaluated daily commencing before the expected day of vaginal patency. 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 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 [14].

Assessment of potential developmental neurotoxicity (cohort 2)

52. *# to be revisited by the DNT group #* 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, on PND 90. The same animals 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. Results of the neurotoxicity assays should be interpreted in relation to appropriate historical control reference ranges.

53. *# to be revisited by the DNT group #* At an appropriate time between PND 49 and PND 56 the cohort 2 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. *# to be revisited by the DNT group #* 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. *# to be revisited by the DNT group #* 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 42-70, 10 male and 10 female cohort 3 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 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 [15, 16]. 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. If the primary antibody response is assayed by counting plaque forming cells, it is permissible to evaluate subgroups of animals on separate days, provided that: subgroup immunization and sacrifice are timed so that PFCs are counted at the peak of the response; that subgroups contain an equal number of male and female offspring from all dose groups, including controls; and that subgroups are evaluated at approximately the same postnatal age. Exposure to the test article will continue until the day before collecting spleens for the PFC response or serum for the ELISA assay.

Assessment of potential reproductive toxicity (*Triggered*)

57. *# to be revisited at the expert group meeting in October 2009 #* 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.*, mechanistic 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 systemic toxicity studies. Such data may be helpful in the planning of the Extended One-Generation Reproductive Toxicity Study as well as in the interpretation of the results. If existing data of suitable duration 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 [17, 18]. In addition, sperm from the cauda epididymis (or vas deferens) is collected using methods that minimise damage for evaluation of sperm motility and morphology [19].

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 [20, 21, 22, 23, 24, 25]. For the evaluation of sperm morphology an epididymal (or vas deferens) sperm sample should be examined as fixed or wet preparations [26] 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 [27]. 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 [28], subsequent analysis may be restricted to control and high-dose males. However, if treatment-related effects are observed, the lower dose groups should also be evaluated.

Gross necropsy

62. At the time of termination or premature death all P and F_1 animals[12] are necropsied and examined macroscopically for any structural abnormalities or pathological changes. 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 oestrous cycle. The uteri of all 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 weight and Tissue Preservation – P and F1 Adult Animals

64. At the time of termination, body weights and wet weights of the organs listed below from all P animals and all F1 adults, from relevant cohorts (as outlined above), are determined as soon as possible after dissection to avoid drying. Unless specified otherwise, paired organs can be weighed individually or combined, consistent with the typical practice of the performing laboratory.

- Uterus (with oviducts and cervix), ovaries
- Testes, epididymides (total and cauda for the samples used for sperm counts)
- Prostate (dorsolateral and ventral part combined). Care should be exercised when trimming the prostate complex to avoid puncture of the fluid filled seminal vesicles. In the event of a treatment-related effect on total prostate weight, the dorsolateral and ventral segments should be carefully dissected after fixation, and weighed separately.
- Seminal vesicles with coagulating glands and their fluids (as one unit)
- Brain, liver, kidneys, heart, spleen, thymus, pituitary, thyroid (post-fixation), adrenal glands and known target organs or tissues.

In addition to the organs listed above, samples of peripheral nerve, muscle, spinal cord, eye plus optic nerve, gastrointestinal tract, urinary bladder, lung, trachea (with thyroid and parathyroid attached), bone marrow, vas deferens (males), and mammary gland (males and females) should be preserved in an appropriate fixative.

[12] **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.

After cohort 1 animals are weighed, the spleen should be divided in half (by length), one half preserved for histopathology, the other half processed for lymphocyte subset analysis (see paragraph 41), and lymph nodes associated with and distant from the route of exposure should be weighed.

At termination, blood from ten randomly selected P males and females and ten randomly selected F1 males and females from cohort 1 is collected for T4 and TSH analysis.

Organ weight and Tissue Preservation – F1 weanlings and juveniles

65. From weanlings subject to gross necropsy on PND 22, the brain, spleen, and thymus should be weighed and retained in an appropriate fixative. Gross abnormalities and target tissues should be saved for possible histological examination. At termination, blood from F1 pups not selected for cohorts is collected for T4 and TSH analysis.

66. *deleted*

67. *deleted*

68. *deleted*

Histopathology – P animals

69. Full histopathology of the organs listed in Paragraph 64 is performed for all high dose and control P animals. Organs demonstrating treatment-related changes should also be examined in 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 oestrus 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.

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 [29]. 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 [30].

Histopathology – F1 animals

Cohort 1 animals

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

All cohort 1B animals have their reproductive and endocrine tissues processed to the block stage. As described in paragraph 30, cohort 1B reproductive and endocrine organs should be examined for histopathology in cases of suspected reproductive or endocrine toxicants. Cohort 1B should also undergo histological examination to further elucidate whether the effects noted in cohort 1A are adverse or adaptive.

73. Ovaries of adult females should contain primordial and growing follicles as well as corpora lutea, therefore, a histopathological examination should be aimed at detecting a qualitative depletion of these parameters in P females and a quantitative evaluation of primordial and small growing follicles, as well as corpora lutea, in F_1 females; the number of animals, ovarian section selection, and section sample size should be statistically appropriate for the evaluation procedure used (although, it is suggested that a minimum of 5% of the ovary be examined). Follicular enumeration may first be conducted on control and high dose animals, and in the event of an adverse effect in the latter, lower doses should be examined. Examination should include enumeration of the number of primordial follicles, which can be combined with small growing follicles, for comparison of treated and control ovaries. Corpora lutea enumeration should be conducted in parallel with oestrous cyclicity testing. 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 animals

75. *# to be revisited by the DNT group #* Neurohistopathology is performed for at least 10 high dose and control cohort 2 animals per sex 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. 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 [31]. 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 should 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 TG 426.

REPORTING

Data

76. Data are reported individually and summarised in tabular form. Where appropriate, for each test group and each generation, the following should be reported: number of animals at the start of the test, number of animals found dead during the test or killed for humane reasons, time of any death or humane kill, number of fertile animals, number of pregnant females, number of females giving birth to a litter, and number of animals showing signs of toxicity. A description of the toxicity, including time of onset, duration, and severity should also be reported.

77. Numerical results should be evaluated by an appropriate, and accepted statistical method. The statistical methods should be selected as part of the study design and 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 and dose-response models cover a broad class of analytical tools that may be appropriate for the data generated under this TG. The report should include sufficient information on the method of analysis and the computer program employed, so that an independent reviewer/statistician can evaluate/re-evaluate 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 and the presence, incidence, and severity of abnormalities, including gross lesions. Target organs, fertility, clinical abnormalities, reproductive and litter performance, body weight changes, mortality and any other toxic and developmental effects should also be assessed. The physico-chemical properties of the test substance, and when available, TK data, including placental transfer and milk excretion, should be taken into consideration when evaluating the test results.

79. *deleted*

Test Report

80. The test report should include the following information obtained in the present study from P, F1 animals and F2 animals (where relevant):

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 (if data are collected at that time)

- 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, except for the period of cohabitation and during lactation), and test material consumption (for dietary/drinking water administration) for P and F_1 animals
- 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
- phenotypic analysis of spleen cells (T-, B-, NK-cells)
- bone marrow cellularity
- 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
- enumeration of corpora lutea in the ovaries of P females
- statistical treatment of results, where appropriate

 # Cohort 2 parameters : will be added after the revision of the DNT battery #

Cohort 3 parameters

- serum IgM antibody titres (sensitization to SRBC or KLH), or splenic IgM PFC units (sensitization to SRBC)

Discussion of results

Conclusions, including NOAEL values for parental and offspring effects

All information not obtained during the study, but useful for the interpretation of the results, should also be provided.

INTERPRETATION OF RESULTS

81. An Extended One-Generation Reproductive 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. Interpretation of the results of the study should take into account all available information on the substance including physico-chemical, TK and toxicodynamic properties, available relevant information on structural analogues, and results of previously-conducted toxicity studies with the test substance (*e.g.*, acute toxicity, toxicity after repeated application, mechanistic studies and studies assessing if there are substantional qualitative and quantative species differences in *in vivo*/*in vitro* metabolic properties). 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 [32].

Cohort 2 spinoff will be included after the revision of the DNT battery

Cohort 3:

Suppression or enhancement of immune function as assessed by TDAR should be evaluated in the context of all observations made. Significance of the outcome of TDAR may be supported by other effects on immunologically related indicators (bone marrow cellularity, weight and histopathology of lymphoid tissues, lymphocyte subset distribution). Effects established by TDAR may be less meaningful in case of other toxicities observed at lower exposure concentrations

83. *Deleted due to differences in legal implications of human testing.*

84. OECD guidance document 43 should be consulted for aid in the interpretation of reproduction and neurotoxicity results [27].

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. Critical Reviews in Toxicology 36, 69-98.

[2] 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.

[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.C., Walls I., Zoetis T. (2005). Direct dosing of preweaning rodents in toxicity testing and research: Deliberations of an ILSI RSI expert working group. International Journal of Toxicology 24, 87-94.

[5] Conolly R.B., Beck B.D., Goodman J.I. (1999). Stimulating research to improve the scientific basis of risk assessment. Toxicological Sciences 49, 1-4.

[6] Takayama S., Akaike M., Kawashima K., Takahashi M., Kurokawa Y. (1995). A Collaborative Study in Japan on Optimal Treatment Period and Parameters for Detection of Male Fertility Disorders Induced by Drugs in Rats. Journal of the American College of Toxicologists. 14, 266-292.

[7] Ulbrich, B., Palmer, A. K. (1995). Detection of effects on male reproduction – a literature survey. Journal of the American College of Toxicologists. 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. Regulatory Toxicology and Pharmacology 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] Goldman J.M., Murr A.S., Buckalew A.R., Ferrell J.M., Cooper R.L. (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.

[12] 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.

[13] Gallavan R.H. Jr, Holson J.F., Stump D.G., Knapp J.F., Reynolds V.L. (1999). Interpreting the toxicologic significance of alterations in anogenital distance: potential for confounding effects of progeny body weights. Reproductive Toxicology 13: 383-390

[14] Korenbrot C.C., Huhtaniemi I.T., Weiner R.I. (1977). Preputial Separation as an External Sign of Pubertal development in the male rat. Biological Reproduction 17, 298-303.

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

[16] 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.

[17] Gray L.E., Ostby J., Ferrell J., Rehnberg G., Linder R., Cooper R., Goldman J., Slott V., Laskey J. (1989). A Dose-Response Analysis of Methoxychlor-Induced Alterations of Reproductive Development and Function in the Rat. Fundamental and Applied Toxicology 12, 92-108.

[18] Robb G.W., Amann R.P., Killian G.J. (1978). Daily Sperm Production and Epididymal Sperm Reserves of Pubertal and Adult Rats. Journal of Reproduction and Fertility 54, 103-107.

[19] Klinefelter G.R., Gray L.E. Jr, Suarez J.D. (1991). The Method of Sperm Collection Significantly Influences Sperm Motion Parameters Following Ethane Dimethanesulfonate Administration in the Rat. Reproductive Toxicology 5, 39-44.

[20] Seed J., Chapin R.E., Clegg E.D., Dostal L.A., Foote R.H., Hurtt M.E., Klinefelter G.R., Makris S.L., Perreault S.D., Schrader S., Seyler D., Sprando R., Treinen K.A., Veeramachaneni D.N., Wise L.D. (1996). Methods for Assessing Sperm Motility, Morphology, and Counts in the Rat, Rabbit, and Dog: a Consensus Report. Reproductive Toxicology 10, 237- 244.

[21] Chapin R.E., Filler R.S., Gulati D., Heindel J.J., Katz D.F., Mebus C.A., Obasaju F., Perreault S.D., Russell S.R., Schrader S. (1992). Methods for Assessing Rat Sperm Motility. Reproductive Toxicology 6, 267-273.

[22] Klinefelter G.R., Roberts N.L., Suarez J.D. (1992). Direct Effects of Ethane Dimethanesulphonate on Epididymal Function in Adult Rats: an *In Vitro* Demonstration**.** Journal of Andrology 13, 409-421.

[23] Slott V.L., Suarez J.D., Perreault S.D. (1991). Rat Sperm Motility Analysis: Methodologic Considerations. Reproductive Toxicology 5, 449-458.

[24] 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.

[25] Toth G.P., Stober J.A., Read E.J., Zenick H., Smith M.K. (1989). The Automated Analysis of Rat Sperm Motility Following Subchronic Epichlorhydrin Administration: Methodologic and Statistical Considerations. Journal of Andrology 10, 401-415.

[26] Linder R.E., Strader L.F., Slott V.L., Suarez J.D. (1992). Endpoints of Spermatoxicity in the Rat After Short Duration Exposures to Fourteen Reproductive Toxicants. Reproductive Toxicology 6, 491-505.

[27] 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

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

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

[30] Lanning L.L., Creasy D.M., Chapin R.E., Mann P.C., Barlow N.J., Regan K.S., Goodman D.G. (2002) Recommended Approaches for the Evaluation of Testicular and Epididymal Toxicity. Toxicologic Pathology 30, 507-520.

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

[32] 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 Chemicals Toxicology 44, 8-16.

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 Reproductive Toxicity Study for Producing a Second Generation (mating of the F_1 offspring to produce F_2 litters)

The Extended One-Generation Reproductive Toxicity 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 Extended One-Generation Reproductive Toxicity 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 Reproductive Toxicity Study 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 Reproductive Toxicity 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 (oestrous 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, oestrus cyclicity was deemed "specific, but not sensitive" after an evaluation of the NTP continuous breeding study database (Chapin *et al.*, 1997). Oestrous cycle length is variable across animals and subject to influence by stress (Matysek, 1989; Roozendaal *et al*., 1995). However, increased oestrous 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 oestrous cycle length that are not associated with overt maternal toxicity would warrant production of a second generation. Oestrous 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 oestrous cycle at necropsy. (The stage of the oestrous 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 e Extended One-Generation Reproductive 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 Reproductive Toxicity 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 Reproductive Toxicity 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 Extended One-Generation Reproductive Toxicity Study 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 Oestrous 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 Oestrous 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 oestrous 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

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. Maria 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 Study

Item	Proposed Addition(s)	Comments and Considerations
1	Oestrus Cyclicity	
	P 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 Oestrous 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 PND 4** 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 parturition	• To calculate precoital interval (pairing to insemination) • To calculate gestational length (insemination to parturition)
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	• Systemic organs (both sexes): brain, CNS,

	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)	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

1
2
3
4

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** PND 4, 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 parturition *	• To calculate precoital interval (pairing to insemination) • To calculate gestational length (insemination to parturition)

1

2

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 assessments
- 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