Adopted:
28 July 2011

OECD GUIDELINE FOR THE TESTING OF CHEMICALS

Extended One-Generation Reproductive Toxicity Study

INTRODUCTION

1. This Test Guideline (TG) 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. 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.
- Cohort 2: assesses the potential impact of chemical exposure on the developing nervous system.
- Cohort 3: assesses the potential impact of chemical exposure on the developing immune system.

2 Decisions on whether to assess the second generation and to omit the developmental neurotoxicity cohort and/or developmental immunotoxicity cohort should reflect existing knowledge for the chemical being evaluated, as well as the needs of various regulatory authorities. The purpose of the Test Guideline is to provide details on how the study can be conducted and to address how each cohort should be evaluated.

3. Procedure for the decision on the internal triggering for producing a 2nd generation is described in Guidance Document 117 (39) for those regulatory authorities using internal triggers.

INITIAL CONSIDERATIONS AND OBJECTIVES

4. 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 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 and when available, information from repeat-dose studies (including screening reproductive toxicity studies, *e.g.* TG 422 (32)), or short term endocrine disrupter screening assays, (*e.g.* Uterotrophic assay - TG 440 (36); and Hershberger assay - TG 441 (37)) is used to detect effects on reproductive organs for males and females. This might include 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 until after sexual maturity (see Guidance Document 151 supporting this Test Guideline (40)).

5. 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. Furthermore, the information obtained from the developmental neurotoxicity and developmental immunotoxicity assessments will 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 for the various endpoints and/or be used to characterize effects detected in previous repeat-dose studies and/or serve as a guide for subsequent testing.

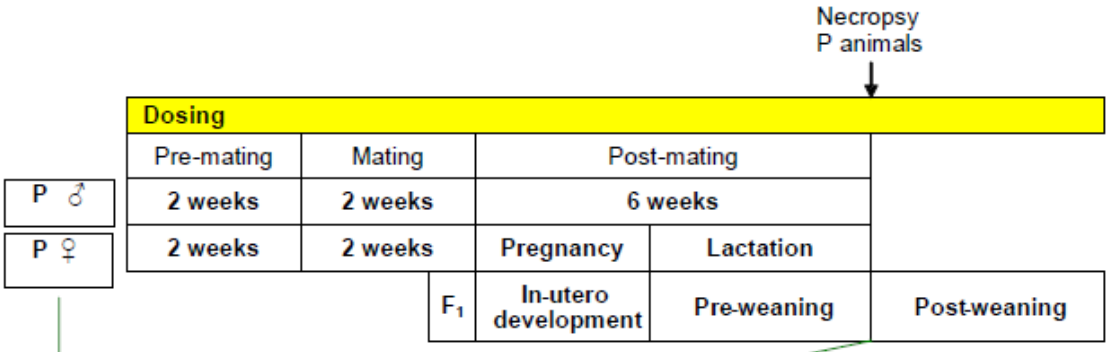
6. A schematic drawing of the protocol is presented in Figure 1. The test substance is administered continuously in graduated doses to several groups of sexually-mature males and females. This parental (P) generation is dosed for a defined pre-mating period (selected based on the available information 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. They should be treated for a minimum of 10 weeks. 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. If a second generation is assessed (see GD 117 (39)), the F_1 offspring will be maintained on treatment until weaning of the F_2, or until termination of the study.

7. 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 33 and 34 and Figure 1) for further investigations, including sexual maturation, reproductive organ integrity and function, neurological and behavioural endpoints, and immune functions.

8. In conducting the study, the guiding principles and considerations outlined in the OECD Guidance Document n° 19 on the recognition, assessment, and use of clinical signs as humane endpoints for experimental animals used in safety evaluations (34) should be followed.

9. When a sufficient number of studies is available to ascertain the impact of this new study design, the Test Guideline will be reviewed and if necessary revised in light of experience gained.

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



Parental generation	Cohort	Designation	Animals/Cohort	Sexual Maturation	Approximate age at necropsy (weeks)
Target is 20 litters per group	1A	Reproductive	20 M +20 F	Yes	13
	1B	Reproductive	20 M +20 F	Yes	14 or 20 if triggered
	2A	Neurotoxicity	10 M +10 F@	Yes	9
	2B	Neurotoxicity	10 M +10 F@	No	3
	3	Immunotoxicity	10 M +10 F@	Yes	8
	Surplus	Spares		No	3

@ one per litter and representative of 20 litters in total where possible

DESCRIPTION OF THE METHOD/PREPARATIONS FOR THE TEST

Animals

Selection of animal species and strain

10. 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, and criteria and recommendations given in this TG refer to this 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

11. 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 (within sex) at initiation of dosing, similar age (approximately 90 days) at mating, and representative of the species and strain under study. 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

12. 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, as a high level of phytoestrogen in the diet might affect some reproductive endpoints. Standardized, open-formula diets in which estrogenic substances have been reduced are recommended (2) (30). 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 each batch of the diet used during the study should be retained under appropriate conditions (*e.g.* frozen at -20°C), until finalisation of the report, in case the results necessitate a further analysis of diet ingredients.

13. 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. Litters are housed with their mothers until weaning. F_1 animals should be housed in small groups of the same sex and treatment group from weaning to termination. If scientifically justified, animals can be housed individually. The level of phytoestrogens contained in the selected bedding material should be minimal.

Number and identification of animals

14. 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.

15. 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 females in each group would have regular (4 or 5-day) oestrous cycles at start of treatment. All F_1 offspring are uniquely identified when neonates are first examined on postnatal day (PND) 0 or 1. Records indicating the litter of origin should be maintained for all F_1 animals, and F_2 animals where applicable, throughout the study.

Test substance

Available information on the test substance

16. 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, toxicokinetics (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, distribution, metabolism and elimination (ADME) and bioaccumulation may be derived from chemical structure, physico-chemical data, extent of plasma protein binding or toxicokinetic (TK) studies, while results from toxicity studies give additional information, *e.g.* on NOAEL, metabolism or induction of metabolism.

Consideration of toxicokinetic data

17. 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 TK data set would be useful in planning the Extended One-Generation Reproductive Toxicity Study:

- Late pregnancy (*e.g.* Gestation Day 20) - maternal blood and foetal blood
- Mid-lactation (PND 10) - maternal blood, pup blood and/or milk
- Early post-weaning (*e.g.* PND 28) - 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

18. 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

19. 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 that may alter its toxic characteristics; and effects on the food or water consumption or the nutritional status of the animals.

Dose selection

20. 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.

21. 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.

22. 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.

23. 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

24. 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

25. 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

26. 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 is therefore 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. At the time of termination, when testicular and epididymal histopathology and analysis of sperm parameters are scheduled, the P and F_1 males, will have been exposed for at least one entire spermatogenic process (6) (7) (8) (9) and GD 151 (40).

27. 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 (see GD 151 (40)).

28. 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 termination after 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 which 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.

29. 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.

30. When the test substance is administered by gavage, the volume of liquid administered at one time should not normally exceed 1 mL/100 g body weight (0.4 mL/100 g body weight is the maximum for oil, *e.g.* corn oil). 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 F_1 animals when administered prior to weaning and during the 2 weeks following weaning. If

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

31. 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.

Litter size

32. 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, *e.g.* 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.

Selection of pups for post-weaning studies (see Figure 1)

33. At weaning (around PND 21) pups from all available litters up to 20 per dose and control group 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 PND 21, the selected F_1 pups are randomly assigned to one of three cohorts of animals, as follows:

Cohort 1 (1A and 1B) = Reproductive/developmental toxicity testing
Cohort 2 (2A and 2B) = Developmental neurotoxicity testing
Cohort 3 = Developmental immunotoxicity testing

Cohort 1A: One male and one female/litter/group (20/sex/group): priority selection for primary assessment of effects upon reproductive systems and of general toxicity.
Cohort 1B: One male and one female/litter/group (20/sex/group): priority selection for follow-up assessment of reproductive performance by mating F_1 animals, when assessed (see GD 117 (39)), and for obtaining additional histopathology data in cases of suspected reproductive or endocrine toxicants, or when results from cohort 1A are equivocal.
Cohort 2A: Total of 20 pups per group (10 males and 10 females per group; one male or one female per litter) assigned for neurobehavioral testing followed by neurohistopathology assessment as adults.
Cohort 2B: Total of 20 pups per group (10 males and 10 females per group; one male or one female per litter) assigned for neurohistopathology assessment at weaning (PND 21 or PND 22). If there are insufficient numbers of animals, preference should be given to assign animals to Cohort 2A.
Cohort 3: Total of 20 pups per group (10 males and 10 females per group; one per litter, where possible). Additional pups may be required from the control group to act as positive control animals in the T-cell dependant antibody response assay (TDAR) at PND 56 ± 3.

34. 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, *e.g.* if a chemical is suspected to be a neurotoxicant, immunotoxicant or reproductive toxicant. These pups may be used for examinations at different timepoints or for the evaluation of supplementary endpoints. Pups not assigned to cohorts will be submitted to clinical biochemistry (paragraph 55) and gross necropsy (paragraph 68).

Second mating of the P animals

35. A second mating is not normally recommended for the P animals, 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

36. 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 and all signs of toxicity are recorded. Twice daily, during the weekend once daily, all animals are observed for severe toxicity, morbidity and mortality.

37. In addition, a more detailed examination of all P and F_1 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

38. 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 (see paragraph 44). 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). All animals are weighed at sacrifice.

39. 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 with selection to a respective cohort.

Oestrous cycles

40. 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-week 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-week smearing period leading into pairing. If the female treatment period is extended in this way (*i.e.* to a 4-week pre-mating treatment) then consideration should be made to purchasing animals youngerand 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 (10) (11).

41. Vaginal smears should be examined daily for all F_1 females in cohort 1A, after the onset of vaginal patency, until the first cornified smear is recorded, in order to determine the time interval between these two events. Oestrous cycles for all F_1 females in cohort 1A should also be monitored for a period of two weeks, commencing around PND 75. In addition, should mating of the F_1 generation be necessary, the vaginal cytology in cohort 1B will be followed from the time of pairing until mating evidence is detected.

Mating and pregnancy

42. In addition to the standard endpoints (*e.g.* 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.

43. 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 also be based on post-coital time to eliminate confounding of postnatal development data, by differences in the duration of pregnancy; however, timing relative to parturition should also be recorded. This is especially important when the test substance exerts an influence on the duration of pregnancy.

Offspring parameters

44. 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, external abnormalities, 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.

45. The anogenital distance (AGD) of each pup should be measured on at least one occasion from PND 0 through 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 (12). The presence of nipples/areolae in male pups should be checked on PND 12 or 13.

46. All selected F_1 animals are evaluated daily for balano-preputial separation or vaginal patency for male/female respectively commencing before the expected day for achievement of these endpoints to detect if sexual maturation occurs early. Any abnormalities of genital organs, such as persistent vaginal thread, hypospadia or cleft penis, should be noted. Sexual maturity of F_1 animals is compared to physical development by determining age and body weight at balano-preputial separation or vaginal opening for male/female respectively (13).

Assessment of potential developmental neurotoxicity (cohorts 2A and 2B)

47. Ten male and 10 female cohort 2A animals and 10 male and 10 female cohort 2B animals, from each treatment group (for each cohort: 1 male or 1 female per litter; all litters represented by at least 1 pup; randomly selected) should be used for neurotoxicity assessments. Cohort 2A animals should be subjected to auditory startle, functional observational battery, motor activity (see paragraphs 48-50), and neuropathology assessments (see paragraphs 74-75). 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 (e.g. intermittent noise), temperature, humidity, lighting, odors, time of day, and environmental distractions. Results of the neurotoxicity assays should be interpreted in relation to appropriate historical control reference ranges. Cohort 2B animals should be used for neuropathology assessment on PND 21 or PND 22 (see paragraphs 74-75).

48. An auditory startle test should be performed on PND 24 (±1 day) using animals in cohort 2A. The day of testing should be counterbalanced across treated and control groups. Each session consists of 50 trials. In performing the auditory startle test, the mean response amplitude on each block of 10 trials (5 blocks of 10 trials) should be determined, with test conditions optimized to produce intra-session habituation. These procedures should be consistent with OECD TG 426 (35).

49. At an appropriate time between PND 63 and PND 75, the cohort 2A animals are subjected to a functional observational battery and an automated test of motor activity. These procedures should be consistent with OECD TGs 424 (33) and 426 (35). The functional observational battery 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. 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. For motor activity, each animal is tested individually. The test session should be long enough to demonstrate intra-session habituation for 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.

50. 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 auditory startle, functional observational battery 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 potential developmental immunotoxicity (cohort 3)

51. At PND 56 (±3 days), 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 in a T-cell dependant antibody response assay, *i.e.* the primary IgM antibody response to a T-cell dependent antigen, such as Sheep Red Blood Cells (SRBC) or Keyhole Limpet Hemocyanin (KLH), consistent with current immunotoxicity testing procedures (14) (15). The response may be evaluated by counting specific plaque-forming cells (PFC) in the spleen or by determining the titer of SRBC- or KLH-specific IgM 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. 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 substance will continue until the day before collecting spleens for the PFC response or serum for the ELISA assay.

Follow-up assessment of potential reproductive toxicity (cohort 1B)

52. Cohort 1B animals can be maintained on treatment beyond PND 90 and bred to obtain a F_2 generation if necessary. Males and females of the same dose group should be cohabited (avoiding the pairing of siblings) for up to two weeks, beginning on or after PND 90, but not exceeding PND 120. Procedures should be similar to those for the P animals. However, based on a weight of evidence, it may suffice to terminate the litters on PND 4 rather than follow them to weaning or beyond.

TERMINAL OBSERVATIONS

Clinical biochemistry / Haematology

53. Systemic effects should be monitored in P animals. 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, clinical biochemistry, assay of T4 and TSH or other examinations suggested by the known effect profile of the test substance (see GD 151 (40)). The following haematological parameters should be examined: haematocrit, haemoglobin concentration, erythrocyte count, total and differential leukocyte count, platelet count and blood clotting time/potential. Investigations of plasma or serum should include: glucose, total cholesterol, urea, creatinine, total protein, albumin and 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. 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 collected a few days before the animals are mated for the second time. Unless existing data from repeated-dose studies indicate that the parameter is not affected by the test substance, urinalysis should be performed prior to termination and the following parameters 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).

54. Systemic effects should also be monitored in F_1 animals. Fasted blood samples from a defined site are taken from ten randomly selected cohort 1A males and females per dose group at termination, stored under appropriate conditions and subjected to standard clinical biochemistry, including the assessment of serum levels for thyroid hormones (T4 and TSH), haematology (total and differential leukocyte plus erythrocyte counts) and urinalysis assessments.

55. The surplus pups at PND 4 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. Blood is also collected for T4 and TSH analysis from weanlings subject to gross necropsy on PND 22 (F_1 pups not selected for cohorts).

Sperm parameters

56. Sperm parameters should be measured in all P generation males unless there is existing data to show that sperm parameters are unaffected in a 90-day study. Examination of sperm parameters should be performed in all cohort 1A males.

57. At termination, testis and epididymis weights are recorded for all P and F_1 (cohort 1A) males. At least one testis and one epididymis are reserved for histopathological examination. The remaining epididymis is used for enumeration of cauda epididymis sperm reserves (16) (17). In addition, sperm from the cauda epididymis (or vas deferens) is collected using methods that minimise damage for evaluation of sperm motility and morphology (18).

58. 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 (19) (20) (21) (22) (23) (24). For the evaluation of sperm morphology, an epididymal (or vas deferens) sperm sample should be examined as fixed or wet preparations (25) 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 (26). Misshapen or large sperm heads may indicate defects in spermiation.

59. If sperm samples are frozen, smears fixed and images for sperm motility analysis recorded at the time of necropsy (27), 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

60. At the time of termination or premature death, all P and F_1 animals 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.

61. 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 and allow correlation with histopathology in reproductive organs. 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 F_1 adult animals

62. At the time of termination, body weights and wet weights of the organs listed below from all P animals and all F_1 adults, from relevant cohorts (as outlined below), are determined as soon as possible after dissection to avoid drying. These organs should then be preserved under appropriate conditions. 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 parts 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.

63. 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), mammary gland (males and females) and vagina should be preserved under appropriate conditions.

64. Cohort 1A animals have all organs weighed and preserved for histopathology.

65. For the investigation of pre- and postnatally induced immunotoxic effects, 10 male and 10 female cohort 1A 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:

- weighing of the lymph nodes associated with and distant from the route of exposure (in addition to the weight of the adrenal glands, the thymus and the spleen, already performed in all cohort 1A animals)
- splenic lymphocyte subpopulation analysis (CD4+ and CD8+ T lymphocytes, B lymphocytes, and natural killer cells) using one half of the spleen, the other half of the spleen being preserved for histopathological evaluation,

Analysis of splenic lymphocyte subpopulations in non-immunized (cohort 1A) animals will determine if exposure is related to a shift in the immunological steady state distribution of "helper" (CD4+) or cytotoxic (CD8+) thymus-derived lymphocytes or natural killer (NK) cells (rapid responses to neoplastic cells and pathogens).

66. Cohort 1B animals should have the following organs weighed and corresponding tissues processed to the block stage:

- Vagina (not weighed)
- Uterus with cervix
- Ovaries
- Testes (at least one)
- Epididymides
- Seminal vesicles and coagulating glands
- Prostate
- Pituitary
- Identified target organs

Histopathology in cohort 1B would be conducted if results from cohort 1A are equivocal or in cases of suspected reproductive or endocrine toxicants.

67. Cohorts 2A and 2B: Developmental neurotoxicity testing (PND 21 or PND 22 and adult offspring). Cohort 2A animals are terminated after behavioural testing, with brain weight recorded and full neurohistopathology for purposes of neurotoxicity assessment. Cohort 2B animals are terminated on PND 21 or PND 22, with brain weight recorded and microscopic examination of the brain for purposes of neurotoxicity assessment. Perfusion fixation is required for cohort 2A animals and optional for cohort 2B animals, as provided in OECD TG 426 (35).

Organ weight and tissue preservation – F_1 weanlings

68. The pups not selected for cohorts, including runts, are terminated after weaning, on PND 22, unless the results indicate the need for further in-life investigations. Terminated pups are subjected to gross necropsy including an assessment of the reproductive organs, as described in paragraphs 62 and 63. For up to 10 pups per sex per group, from as many litters as possible, brain, spleen, and thymus should be weighed and retained under appropriate conditions. In addition, mammary tissues for these male and female pups may be preserved for further microscopic analysis[1] (see GD 151 (40)). Gross abnormalities and target tissues should be saved for possible histological examination.

Histopathology – P animals

69. Full histopathology of the organs listed in paragraphs 62 and 63 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

[1] Research has shown the mammary gland, especially in early life mammary gland development, to be a sensitive endpoint for oestrogen action. It is recommended that endpoints involving pup mammary glands of both sexes be included in this Test Guideline, when validated.

offspring, or for which oestrous cyclicity or sperm number, motility, or morphology were affected, and all gross lesions should be subjected to histopathological evaluation.

Histopathology – F_1 animals

Cohort 1 animals

70. Full histopathology of the organs listed in paragraphs 62 and 63 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. For the evaluation of pre- and postnatally induced effects on lymphoid organs also the histopathology on the collected lymph nodes and bone marrow should be evaluated of 10 male and 10 female cohort 1A animals next to histopathological evaluation of the thymus, spleen, and the adrenal glands already performed in all 1A animals.

71. Reproductive and endocrine tissues from all cohort 1B animals, processed to the block stage as described in paragraph 66, should be examined for histopathology in cases of suspected reproductive or endocrine toxicants. Cohort 1B should also undergo histological examination if results from cohort 1A are equivocal.

72. 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 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. 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 (see GD 151 (40)). Corpora lutea assessment should be conducted in parallel with oestrous cyclicity testing so that the stage of the cycle can be taken into account in the assessment. Oviduct, uterus and vagina are examined for appropriate organ-typic development.

73. 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 (38). When possible, sections of the rete testis should be examined. 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

74. Neurohistopathology is performed for all high-dose and control cohort 2A animals per sex following completion of neurobehavioral testing (after PND 75, but not to exceed PND 90). Brain histopathology is performed for all high-dose and control cohort 2B animals per sex on PND 21 or PND 22. Organs or tissues demonstrating treatment-related changes should also be examined for the animals in the lower dose groups to aid in determining a NOAEL. For cohort 2A and 2B animals, multiple sections are examined from the brain to allow examination of olfactory bulbs, cerebral cortex, hippocampus, basal ganglia, thalamus, hypothalamus, mid-brain (thecum, tegmentum, and cerebral peduncles), brain-stem and cerebellum. For cohort 2A only, the eyes (retina and optic nerve) and samples of peripheral nerve, muscle and spinal cord are examined. All neurohistological procedures should be consistent with OECD TG 426 (35).

75. Morphometric (quantitative) evaluations should be performed on representative areas of the brain (homologous sections carefully selected based on reliable microscopic landmarks) and may include linear and/or areal measurements of specific brain regions. At least three consecutive sections should be taken at each landmark (level) in order to select the most homologous and representative section for the specific brain area to be evaluated. 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 (28). 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 tissue fixation, 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. When 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 associated with prolonged storage in fixative.

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), 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. Special attention should be given to sex-specific changes. 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.

Test report

79. The test report should include the following information obtained in the present study from P, F_1 animals and F_2 animals (where relevant):

Test substance:

- All relevant available information on the substance, toxicokinetic and toxicodynamic properties of the test substance;
- 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 (*e.g.* diet, drinking water), 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);
- Detailed description of the randomization procedures to select pups for culling and to assign pups to test groups;
- Environmental conditions;
- List of study personnel, including professional training;

Results (summary and individual data by sex and dose):

- 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;

- Number of P and F_1 females with normal or abnormal oestrous 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;
- 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 F_1 females;
- Statistical treatment of results, where appropriate;

Cohort 2 parameters:

- Detailed description of the procedures used to standardize observations and procedures as well as operational definitions for scoring observations;
- List of all test procedures used, and justification for their use;
- Details of the behavioural/functional, neuropathological and morphometric procedures used, including information and details on automated devices;
- Procedures for calibrating and ensuring the equivalence of devices and the balancing of treatment groups in testing procedures;
- Short justification explaining any decisions involving professional judgment;
- Detailed description of all behavioural/functional, neuropathological and morphometric findings by sex and dose group, including both increases and decreases from controls;
- Brain weight;
- Any diagnoses derived from neurological signs and lesions, including naturally-occurring diseases or conditions;
- Images of exemplar findings;
- Low-power images to assess homology of sections used for morphometry;
- Statistical treatment of results, including statistical models used to analyze the data, and the results, regardless of whether they were significant or not;
- Relationship of any other toxic effects to a conclusion about the neurotoxic potential of the test chemical, by sex and dose group;
- Impact of any toxicokinetic information on the conclusions;

- Data supporting the reliability and sensitivity of the test method (*i.e.* positive and historical control data);
- Relationships, if any, between neuropathological and functional effects;
- NOAEL or benchmark dose for dams and offspring, by sex and dose group;
- Discussion of the overall interpretation of the data based on the results, including a conclusion of whether or not the chemical caused developmental neurotoxicity and the NOAEL;

Cohort 3 parameters

- Serum IgM antibody titres (sensitization to SRBC or KLH), or splenic IgM PFC units (sensitization to SRBC);
- Performance of the TDAR method should be confirmed as part of the optimisation process by laboratory setting up the assay for the first time, and periodically (*e.g.* yearly) by all laboratories;
- Discussion of the overall interpretation of the data based on the results, including a conclusion of whether or not the chemical caused developmental immunotoxicity and the NOAEL;

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 (*e.g.* similarities of effects to any known neurotoxicants), should also be provided.

INTERPRETATION OF RESULTS

80. 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, as necessary. In particular, the study provides information on the reproductive system, and on development, growth, survival, and functional endpoints of offspring up to PND 90.

81. 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 substantial qualitative and quantitative 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 (29).

Cohort 2 (Developmental neurotoxicity)

82. Neurobehavioral and neuropathology results should be interpreted in the context of all findings, using a weight-of-evidence approach with expert judgment. Patterns of behavioural or morphological findings, if present, as well as evidence of dose-response should be discussed. The evaluation of developmental neurotoxicity, including human epidemiological studies or case reports, and experimental animal studies (*e.g.* toxicokinetic data, structure-activity information, data from other toxicity studies) should be included in this characterization. Evaluation of data should include a discussion of both the biological and statistical significance. The evaluation should include the relationship, if any, between observed neuropathological and behavioural alterations. For guidance on the interpretation of developmental neurotoxicity results, refer to OECD TG 426 (35) and Tyl et al., 2008 (31).

Cohort 3 (Developmental immunotoxicity)

83. Suppression or enhancement of immune function as assessed by TDAR (T-cell dependent antibody response), 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 (*e.g.* 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.

84. OECD Guidance Document 43 should be consulted for aid in the interpretation of reproduction and neurotoxicity results (26).

LITERATURE

(1) Cooper, R.L., J.C. Lamb, S.M. Barlow, K. Bentley, A.M. Brady, N. Doerr, D.L. Eisenbrandt, P.A. Fenner-Crisp, R.N. Hines, L.F.H. Irvine, C.A. Kimmel, H. Koeter, A.A. Li, S.L. Makris, L.P. Sheets, G.J.A. Speijers and K.E. Whitby (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. and I. Walls (2003), *Principles and Practices for Direct Dosing of Pre-Weaning Mammals in Toxicity Testing and Research*, ILSI Press, Washington, DC.

(4) Moser, V.C., I. Walls and T. Zoetis (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., B.D. Beck, and J.I. Goodman (1999), "Stimulating Research to Improve the Scientific Basis of Risk Assessment", *Toxicological Sciences*, 49, 1-4.

(6) Ulbrich, B. and A.K. Palmer (1995), "Detection of Effects on Male Reproduction – a Literature Survey", *Journal of the American College of Toxicologists*, 14, 293-327.

(7) Mangelsdorf, I., J. Buschmann and B. Orthen (2003), "Some Aspects Relating to the Evaluation of the Effects of Chemicals on Male Fertility", *Regulatory Toxicology and Pharmacology*, 37, 356-369.

(8) Sakai, T., M. Takahashi, K. Mitsumori, K. Yasuhara, K. Kawashima, H. Mayahara and Y. Ohno (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.

(9) 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.

(10) Goldman, J.M., A.S. Murr, A.R. Buckalew, J.M. Ferrell and R.L. Cooper (2007), "The Rodent Estrous Cycle: Characterization of Vaginal Cytology and its Utility in Toxicological Studies", *Birth Defects Research*, Part B, 80 (2), 84-97.

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

(12) Gallavan, R.H. Jr, J.F. Holson, D.G. Stump, J.F. Knapp and V.L. Reynolds (1999), "Interpreting the Toxicologic Significance of Alterations in Anogenital Distance: Potential for Confounding Effects of Progeny Body Weights", *Reproductive Toxicology*, 13: 383-390.

(13) Korenbrot, C.C., I.T. Huhtaniemi and R.I. Weiner (1977), "Preputial Separation as an External Sign of Pubertal Development in the Male Rat", *Biological Reproduction*, 17, 298-303.

(14) Ladics, G.S. (2007), "Use of SRBC Antibody Responses for Immunotoxicity Testing", *Methods*, 41, 9-19.

(15) Gore, E.R., J. Gower, E. Kurali, J.L. Sui, J. Bynum, D. Ennulat and D.J. Herzyk (2004), "Primary Antibody Response to Keyhole Limpet Hemocyanin in Rat as a Model for Immunotoxicity Evaluation", *Toxicology*, 197, 23-35.

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

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

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

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

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

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

(22) Slott, V.L., J.D. Suarez and S.D. Perreault (1991), "Rat Sperm Motility Analysis: Methodologic Considerations", *Reproductive Toxicology*, 5, 449-458.

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

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

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

(26) OECD (2008), *Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment*, Series on Testing and Assessment, No. 43, ENV/JM/MONO(2008)16, OECD, Paris.

(27) Working, P.K., M. Hurtt (1987), "Computerized Videomicrographic Analysis of Rat Sperm Motility", *Journal of Andrology*, 8, 330-337.

(28) 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.

(29) Stütz, N., B. Bongiovanni, M. Rassetto, A. Ferri, A.M. Evangelista de Duffard, and R. Duffard (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.

(30) Thigpen, JE, K.D.R. Setchell, J.K. Haseman, H.E. Saunders, G.F. Caviness, G.E. Kissling, M.G. Grant and D.B. Forsythe (2007), "Variations in Phytoestrogen Content between Different Mill Dates of the Same Diet Produces Significant Differences in the Time of Vaginal Opening in CD-1 Mice and F344 Rats but not in CD Sprague Dawley Rats", *Environmental health perspectives*, 115(12), 1717-1726.

(31) Tyl, R.W., K. Crofton, A. Moretto, V. Moser, L.P. Sheets and T.J. Sobotka (2008), "Identification and Interpretation of Developmental Neurotoxicity Effects: a Report from the ILSI Research Foundation/Risk Science Institute Expert Working Group on Neurodevelopmental Endpoints", *Neurotoxicology and Teratology*, 30: 349-381.

(32) OECD (1996), *Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test*, OECD Guideline for Testing of Chemicals, No. 422, OECD, Paris.

(33) OECD (1997), *Neurotoxicity Study in Rodents*, OECD Guideline for Testing of Chemicals, No.424, OECD, Paris.

(34) OECD (2000), *Guidance Document on the recognition, assessment, and use of clinical signs as humane endpoints for experimental animals used in safety evaluations*, Series on Testing and Assessment, No. 19, ENV/JM/MONO(2000)7, OECD, Paris.

(35) OECD (2007), *Developmental Neurotoxicity Study*, OECD Guideline for Testing of Chemicals, No. 426, OECD, Paris.

(36) OECD (2007), *Uterotrophic Bioassay in Rodents: A short-term Screening Test for Oestrogenic Properties*, OECD Guideline for Testing of Chemicals, No. 440, OECD, Paris.

(37) OECD (2009), *Hershberger Bioassay in Rats: A Short-term Screening Assay for (Anti)Androgenic Properties*, OECD Guideline for Testing of Chemicals, No. 441, OECD, Paris.

(38) OECD (2009), *Guidance Document for Histologic Evalution of Endocrine and Reproductive Test in Rodents*, Series on Testing and Assessment, No. 106, OECD, Paris.

(39) OECD (2011), *Guidance Document on the Current Implementation of Internal Triggers in the Extended One Generation Reproductive Toxicity Study in the United States and Canada*, Series on Testing and Assessment, No. 117, ENV/JM/MONO(2011)21, OECD, Paris.

(40) OECD (2011), *Guidance Document supporting TG 443: Extended One Generation Reproductive Toxicity Study*, Series on Testing and Assessment, No. 151, OECD, Paris (not yet available).

Appendix A

Measures and Observations Included in the Functional Observational Battery (Cohort 2A)

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	
Gait Abnormalities	Landing Foot Splay	
Arousal	Forelimb Grip Strength	
Stereotypy	Hindlimb Grip Strength	
Bizarre Behaviour		
Stains		
Respiratory Abnormalities		