Adopted:
7 September 2009

OECD GUIDELINE FOR THE TESTING OF CHEMICALS

The Amphibian Metamorphosis Assay

INTRODUCTION

1. The need to develop and validate an assay capable of detecting substances active in the thyroid system of vertebrate species originates from concerns that environmental levels of chemicals may cause adverse effects in both humans and wildlife. In 1998, the OECD initiated a high-priority activity to revise existing guidelines and to develop new guidelines for the screening and testing of potential endocrine disrupters. One element of the activity was to develop a test guideline for the screening of substances active on the thyroid system of vertebrate species. Both an enhancement of the 28-day repeat dose in rodents (Test Guideline 407) and the Amphibian Metamorphosis Assay (AMA) were proposed. The enhanced TG 407 underwent validation and a revised Test guideline has been issued. The Amphibian Metamorphosis Assay (AMA) underwent an extensive validation programme which included intra- and inter-laboratory studies demonstrating the relevance and reliability of the assay (1, 2). Subsequently, the validation of the assay was subject to peer-review by a panel of independent experts (3). This Test Guideline is the outcome of the experience gained during the validation studies for the detection of thyroid active substances, and of work conducted elsewhere in OECD member countries.

PRINCIPLE OF THE TEST

2. The Amphibian Metamorphosis Assay (AMA) is a screening assay intended to empirically identify substances which may interfere with the normal function of the hypothalamic-pituitary-thyroid (HPT) axis. The AMA represents a generalized vertebrate model to the extent that it is based on the conserved structures and functions of the HPT axis. It is an important assay because amphibian metamorphosis provides a well-studied, thyroid-dependent process which responds to substances active within the HPT axis, and it is the only existing assay that detects thyroid activity in an animal undergoing morphological development.

3. The general experimental design entails exposing stage 51 *Xenopus laevis* tadpoles to a minimum of three different concentrations of a test chemical and a dilution water control for 21 days. There are four replicates of each test treatment. Larval density at test initiation is 20 tadpoles per test tank for all treatment groups. The observational endpoints are hind limb length, snout to vent length (SVL), developmental stage, wet weight, thyroid histology, and daily observations of mortality.

DESCRIPTION OF THE METHOD

Test Species

4. *Xenopus laevis* is routinely cultured in laboratories worldwide and is easily obtainable through commercial suppliers. Reproduction can be easily induced in this species throughout the year using human chorionic gonadotropin (hCG) injections and the resultant larvae can be routinely reared to selected developmental stages in large numbers to permit the use of stage-specific test protocols. It is preferred that larvae used in the assay are derived from in-house adults. As an alternative although this is not the

preferred procedure, eggs or embryos may be shipped to the laboratory performing the test and allowed to acclimate; the shipping of larval stages for use in the test is unacceptable.

Equipment and Supplies

5. The following equipment and supplies are needed for the conduct of this assay:

a. Exposure system (see description below);
b. Glass or stainless steel aquaria (see description below);
c. Breeding tanks
d. Temperature controlling apparatus (e.g., heaters or coolers (adjustable to 22º ± 1°C));
e. Thermometer;
f. Binocular dissection microscope;
g. Digital camera with at least 4 megapixel resolution and micro function;
h. Image digitizing software;
i. Petri dish (e.g. 100 x 15 mm) or transparent plastic chamber of comparable size;
j. Analytical balance capable of measuring to 3 decimal places (mg);
k. Dissolved oxygen meter;
l. pH meter;
m. Light intensity meter capable of measuring in lux units;
n. Miscellaneous laboratory glassware and tools;
o. Adjustable pipetters (10 to 5,000 µL) or assorted pipettes of equivalent sizes.
p. Test chemical in sufficient quantities to conduct the study, preferably of one lot #
q. Analytical instrumentation appropriate for the chemical on test or contracted analytical services

Chemical Testability

6. The AMA is based upon an aqueous exposure protocol whereby test chemical is introduced into the test chambers via a flow through system. Flow-through methods however, introduce constraints on the types of chemicals that can be tested, as determined by the physicochemical properties of the compound. Therefore, prior to using this protocol, baseline information about the chemical should be obtained that is relevant to determining the testability, and the OECD Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures should be consulted (4). Characteristics which indicate that the chemical may be difficult to test in aquatic systems include: high octanol water partitioning coefficients (log K_{ow}), high volatility, susceptibility to hydrolysis, and susceptibility to photolysis under ambient laboratory lighting conditions. Other factors may also be relevant to determining testability and should be determined on a case by case basis. If a successful test is not possible for the chemical using a flow-through test system, a static renewal system may be employed. If neither system is capable of accommodating the test chemical, then the default is to not test it using this protocol.

Exposure System

7. A flow-through diluter system is preferred, when possible, over a static renewal system. If physical and/or chemical properties of any of the test substances are not amenable to a flow-through diluter system, then an alternative exposure system (e.g., static-renewal) can be employed. The system components should have water-contact components of glass, stainless steel, and/or Teflon®. However, suitable plastics can be utilized if they do not compromise the study. Exposure tanks should be glass or stainless steel aquaria, equipped with standpipes that result in an approximate tank volume between 4.0 and 10.0 L and minimum water depth of 10 to 15 cm. The system should be capable of supporting all exposure concentrations and a control, with four replicates per treatment. The flow rate to each tank should constant

in consideration of both the maintenance of biological conditions and chemical exposure (e.g. 25 mL/min). The treatment tanks should be randomly assigned to a position in the exposure system in order to reduce potential positional effects, including slight variations in temperature, light intensity, etc. Fluorescent lighting should be used to provide a photoperiod of 12 hr light: 12 hr dark at an intensity that ranges from 600 to 2,000 lux (lumens/m2) at the water surface. Water temperature should be maintained at 22º ± 1°C, pH maintained between 6.5 to 8.5, and the dissolved oxygen (DO) concentration > 3.5 mg/L (> 40% of the air saturation) in each test tank. As a minimum water temperature, pH and dissolved oxygen should be measured weekly; temperature should preferably be measured continuously in at least one test vessel. Annex 1 outlines the experimental conditions under which the protocol should be executed. For further information on setting up flow-through exposure systems and/or static renewal systems, please refer to the ASTM Standard Guide for Conducting Acute Toxicity Tests on Test Materials with Fishes, Macroinvertebrates, and Amphibians (5) and general aquatic toxicology tests.

Water quality

8. Any water that is locally available (e.g. springwater or charcoal-filtered tap water) and permits normal growth and development of *X. laevis* tadpoles could be used. Because local water quality can differ substantially from one area to another, analysis of water quality should be undertaken, particularly, if historical data on the utility of the water for raising *Xenopus* is not available. Special attention should be given that the water is free of copper, chlorine and chloramines, all of which are toxic to frogs and tadpoles. It is further recommended to analyze the water concerning background levels of fluoride, perchlorate and chlorate (by-product of drinking water disinfection) as all of these anions are substrates of the iodine transporter of the thyroid gland and elevated levels of each of these anions may confound the study outcome. Analysis should be performed before testing begins and the testing water should normally be free from these anions.

Iodide Concentration in Test Water

9. In order for the thyroid gland to synthesize TH, sufficient iodide needs to be available to the larvae through a combination of aqueous and dietary sources. Currently, there are no empirically derived guidelines for minimal iodide concentrations. However, iodide availability may affect the responsiveness of the thyroid system to thyroid active agents and is known to modulate the basal activity of the thyroid gland, an aspect that deserves attention when interpreting the results from thyroid histopathology. Therefore, measured aqueous iodide concentrations from the test water should be reported. Based on the available data from the validation studies, the protocol has been demonstrated to work well when test water iodide (I^-) concentrations ranged between 0.5 and 10µg/L. Ideally, the minimum iodide concentration in the test water should be 0.5µg/L. If the test water is reconstituted from deionized water, iodine should be added at a minimum concentration of 0.5µg/L. Any additional supplementation of the test water with iodine or other salts should be noted in the report.

Holding of animals

Adult Care and Breeding

10. Adult care and breeding is conducted in accordance with standard guidelines (6) and the reader is directed to the standard guide for performing FETAX for more detailed information. Such standard guidelines provide an example of appropriate care and breeding methods, but strict adherence is not required. To induce breeding, pairs (3-5) of adult females and males are injected with human chorionic gonadotropin (hCG). Female and male specimens are injected with approximately 800 IU-1000IU and 600 IU-800 IU, respectively, of hCG dissolved in 0.6-0.9% saline solution. Breeding pairs are held in large tanks, undisturbed and under static conditions in order to promote amplexus. The bottom of each breeding

tank should have a false bottom of stainless steel or plastic mesh which permits the egg masses to fall to the bottom of the tank. Frogs injected in the late afternoon will usually deposit most of their eggs by mid morning of the next day. After a sufficient quantity of eggs are released and fertilized, adults should be removed from the breeding tanks.

Larval Care and Selection

11. After the adults are removed from the breeding tanks, the eggs are collected and evaluated for viability using a representative sub-set of the embryos from all breeding tanks. The best individual spawn(s) (2-3 recommended to evaluate the quality of the spawns) should be retained based upon embryo viability and the presence of an adequate number (minimum of 1500) of embryos. All the organisms used in a study should originate from a single spawning event (i.e., the spawns should not be co-mixed). The embryos are transferred into a large flat pan or dish and all obvious dead or abnormal eggs (see definition in (5))are removed using a pipette or eyedropper. The sound embryos from each of the three spawns are transferred into three separate hatching tanks. Four days after being placed in the hatching tanks, the best spawn, based on viability and hatching success, is selected and the larvae are transferred into an appropriate number of rearing tanks at 22° ±1°C. In addition, some additional larvae are moved into extra tanks for use as replacements in the event that mortalities occur in the rearing tanks during the first week. This procedure maintains consistent organism density and thereby reduces developmental divergence within the cohort of a single spawn. All rearing tanks should be siphoned clean daily. As a precaution, vinyl or nitrile gloves are preferred to latex gloves. Mortalities should be removed daily and replacement larvae should be added back to maintain the organism density during the first week. Feeding should occur at least twice per day.

12. During the pre-exposure phase, tadpoles are acclimated to the conditions of the actual exposure phase including; the type of food, temperature, light-dark cycle, and the culture medium. Therefore, it is recommended that the same culture/dilution water be used during the pre-exposure phase and the exposure phase. If a static culture system is used for maintaining tadpoles during the pre-exposure phase, the culture medium should be replaced completely at least twice per week. Crowding, caused by high larval densities during the pre-exposure period, should be avoided because such effects could markedly affect tadpole development during the subsequent testing phase. Therefore, the rearing density should not exceed approximately four tadpoles/L culture medium (static exposure system) or 10 tadpoles/L culture medium (with e.g. 50 mL/min flow rate in the pre-exposure or culturing system). Under these conditions, tadpoles should develop from stages 45/46 to stage 51 within twelve days. Representative tadpoles of this stock population should be inspected daily for developmental stage in order to estimate the appropriate time point for initiation of exposure. Care should be used to minimize stress and trauma to the tadpoles, especially during movement, cleaning of aquaria, and manipulation of larvae. Stressful conditions/activities should be avoided such as loud and/or incessant noise, tapping on aquaria, vibrations in the aquaria, excessive activity in the laboratory, and rapid changes in environmental media (light availability, temperature, pH, DO, water flow rates, etc.) If tadpoles do not develop to stage 51 within 17 days after fertilization, excessive stress should be considered as a potential culprit.

Larval Culture and Feeding

13. Tadpoles are fed with e.g. Sera Micron® (Sera GmbH, Heinsberg, Germany) throughout the pre-exposure period (after Nieuwkoop and Faber (NF) stage 45/46) and during the entire test period of 21 days, or other diet that has demonstrated to allow equal performance of the Amphibian Metamorphosis Assay. The feeding regime during the pre-exposure period should be carefully adjusted to meet the demands of the developing tadpoles. That is, small portions of food should be provided to the newly hatched tadpoles several times per day (at least twice). Excess food should be avoided in order *i)* to maintain water quality and *ii)* to prevent the clogging of gill filters with food particles and detritus. For Sera Micron ®, the daily

food rations should be increased along with tadpole growth to approximately 30 mg/animal/day shortly before test initiation. Sera Micron®, a commercially available tadpole food that has been shown in the validation studies to support proper growth and development of *X. laevis* tadpoles, is a fine particulate that stays suspended in the water column for a long period of time and is subject to washing out with the flow. Therefore, the total daily amount of food should be divided into smaller portions and fed at least twice daily. For Sera Micron ®, the feeding regime is outlined in Table 1**.** Feeding rates should be recorded. Sera Micron® can be fed dry or as a stock solution prepared in dilution water. Such a stock solution should be freshly prepared every other day and stored at 4º C when not in use.

Table 1. Feeding regime for X. laevis tadpoles during the in-life portion of the AMA in flow-through conditions

Study Day	Food ration (mg Sera Micron®/animal/day)
0-4	30
5-7	40
8-10	50
11-14	70
15-21	80

Analytical Chemistry

14. Prior to conducting a study, the stability of the test compound should be evaluated using existing information on its solubility, degradability, and volatility Test solutions from each replicate tank at each concentration should be sampled for analytical chemistry analyses at test initiation (day 0), and weekly during the test for a minimum of four samples. It is also recommended that each test concentration be analyzed during system preparation, prior to test initiation, to verify system performance. In addition, it is recommended that stock solutions be analyzed when they are changed, especially if the volume of the stock solution does not provide adequate amounts of chemical to span the duration of routine sampling periods. In the case of chemicals which cannot be detected at some or all of the concentrations used in a test, stock solutions should be measured and system flow rates recorded in order to calculate nominal concentrations. .

Chemical Delivery

15. The method used to introduce the test chemical to the system can vary depending on its physicochemical properties. Water soluble compounds can be dissolved in aliquots of test water at a concentration which allows delivery at the target test concentration in a flow through system. Chemicals which are liquid at room temperature and sparingly soluble in water can be introduced using liquid:liquid saturator methods. Chemicals which are solid at room temperature and are sparingly soluble in water can be introduced using glass wool column saturators (7). The preference is to use a carrier-free test system, however different test chemicals will possess varied physicochemical properties that will likely require different approaches for preparation of chemical exposure water. It is preferred that effort be made to avoid solvents or carriers because: *i)* certain solvents themselves may result in toxicity and/or undesirable or unexpected endocrinological responses, *ii)* testing chemicals above their water solubility (as can frequently occur through the use of solvents) can result in inaccurate determinations of effective concentrations, and *iii)* the use of solvents in longer-term tests can result in a significant degree of "biofilming" associated with microbial activity. For difficult to test substances, a solvent may be employed as a last resort, and the OECD Guidance Document on aquatic toxicity testing of difficult substances and mixtures should be consulted (4) to determine the best method. The choice of solvent will be determined by the chemical properties of the substance. Solvents which have been found to be effective for aquatic toxicity testing include acetone, ethanol, methanol, dimethyl formamide and triethylene glycol.

In case a solvent carrier is used, solvent concentrations should be below the chronic No Observed Effect Concentration (NOEC); the OECD Guidance Document recommends a maximum of 100µl/L; a recent review recommends that solvent concentrations as low as 20µl/L of dilution water be used (12). If solvent carriers are used, appropriate solvent controls (clean water) should be evaluated in addition to non-solvent controls. If it is not possible to administer a chemical via the water, either because of physicochemical characteristics (low solubility) or limited chemical availability, introducing it via the diet may be considered. Preliminary work has been conducted on dietary exposures; however, this route of exposure is not commonly used. The choice of method should be documented and analytically verified.

Selection of test concentrations

Establishing the High Test Concentration

16. For the purposes of this test, the high test concentration should be set by the solubility limit of the test substance; the maximum tolerated concentration (MTC) for acutely toxic chemicals; or 100 mg/L, whichever is lowest.

17. The MTC is defined as the highest test concentration of the chemical which results in less than 10% acute mortality. Using this approach assumes that there are existing empirical acute mortality data from which the MTC can be estimated. Estimating the MTC can be inexact and typically requires some professional judgment. Although the use of regression models may be the most technically sound approach to estimating the MTC, a useful approximation of the MTC can be derived from existing acute data by using 1/3 of the acute LC_{50} value. However, acute toxicity data may be lacking for the species on test. If species specific acute toxicity data are not available, then a 96-hour LC_{50} test can be completed with tadpoles that are representative (i.e., same stage) of those on test in the AMA. Optionally, if data from other aquatic species are available (e.g. LC_{50} studies in fish or other amphibian species), then professional judgment may be used to estimate a likely MTC based on inter-species extrapolation.

18. Alternatively, if the chemical is not acutely toxic and is soluble above 100 mg/L, then 100 mg/L should be considered the highest test concentration (HTC), as this concentration is typically considered "practically non-toxic."

19. Although not the recommended procedure, static renewal methods may be used where flow-through methods are inadequate to achieve the MTC. If static renewal methods are used, then the stability of the test chemical concentration should be documented and remain within the performance criteria limits. Twenty-four hour renewal periods are recommended. Renewal periods exceeding 72 hours are not acceptable. Additionally, water quality parameters (e.g. DO, temperature, pH, etc.) should be measured at the end of each renewal period, immediately prior to renewal.

Test Concentration Range

20. There is a required *minimum* of three test concentrations and a clean water control (and vehicle control if necessary). The minimum test concentration differential between the highest and lowest should be about one order of magnitude. The maximum dose separation is 0.1 and the minimum is 0.33.

PROCEDURE

Test Initiation and Conduct

Day 0

21. The exposure should be initiated when a sufficient number of tadpoles in the pre-exposure stock population have reached developmental stage 51, according to Nieuwkoop and Faber (8), and which are less than or equal to 17 days of age post fertilization. For selection of test animals, healthy and normal looking tadpoles of the stock population should be pooled in a single vessel containing an appropriate volume of dilution water. For developmental stage determination, tadpoles should be individually removed from the pooling tank using a small net or strainer and transferred to a transparent measurement chamber (e.g., 100 mm Petri dish) containing dilution water. For stage determination, it is preferred not to use anesthesia, however one may individually anesthetize the tadpoles using 100 mg/L tricaine methanesulfonate (e.g. MS-222), appropriately buffered with sodium bicarbonate (pH 7.0), before handling. If used, methodology for appropriately using e.g. MS-222 for anesthesia should be obtained from experienced laboratories and reported with the test results. Animals should be carefully handled during this transfer in order to minimize handling stress and to avoid any injury.

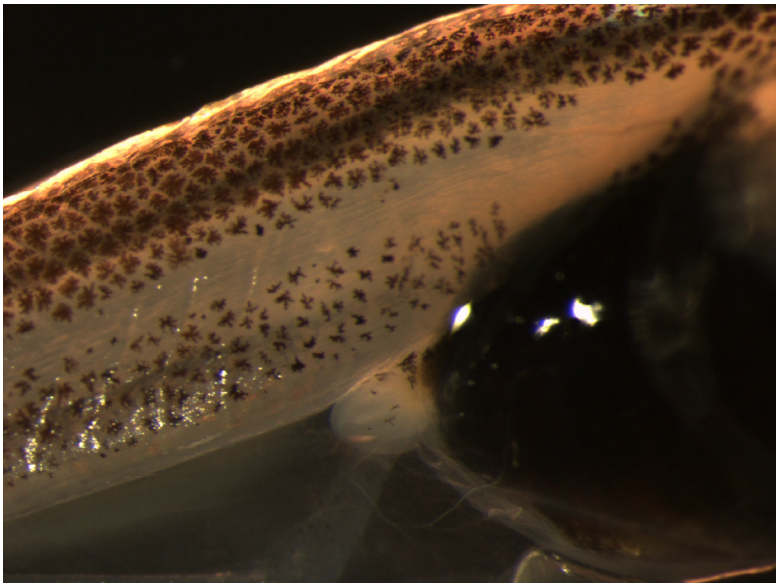
22. The developmental stage of the animals is determined using a binocular dissection microscope. To reduce the ultimate variability in developmental stage, it is important that this staging be conducted as accurately as possible. According to Nieuwkoop and Faber (8), the primary developmental landmark for selecting stage 51 organisms is hind limb morphology. The morphological characteristics of the hind limbs should be examined under the microscope. While the complete Nieuwkoop and Faber (8) guide should be consulted for comprehensive information on staging tadpoles, one can reliably determine stage using prominent morphological landmarks. The following table can be used to simplify and standardize the staging process throughout the study by identifying those prominent morphological landmarks associated with different stages, assuming that development is normal.

Table 2. Prominent morphological staging landmarks based on Neuwkoop and Faber guidance.

Prominent Morphological Landmarks	Developmental Stage															
	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66
Hindlimb	X	X	X	X	X	X	X									
Forelimb						X	X	X	X	X						
Craniofacial structure										X	X	X	X			
Olfactory nerve morphology											X	X	X			
Tail length													X	X	X	X

23. For test initiation, all tadpoles should be at stage 51. The most prominent morphological staging landmark for that stage is hind limb morphology, which is demonstrated in Figure 1.

Figure 1. Hind limb morphology of a stage 51 *X. laevis* tadpole.



24. In addition to the developmental stage selection, an optional size selection of the experimental animals may be used. For this purpose, the whole body length (not SVL) should be measured at day 0 for a sub-sample of approximately 20 NF stage 51 tadpoles. After calculation of the mean whole body length for this group of animals, minimum and maximum limits for the whole body length of experimental animals can be set by allowing a range of the mean value ± 3 mm (mean values of whole body length range between 24.0 and 28.1 mm for stage 51 tadpoles). However, developmental staging is the primary parameter in determining the readiness of each test animal. Tadpoles exhibiting grossly visible malformations or injuries should be excluded from the assay.

25. Tadpoles that meet the stage criteria described above are held in a tank of clean culture water until the staging process is completed. Once the staging is completed, the larvae are randomly distributed to exposure treatment tanks until each tank contains 20 larvae. Each treatment tank is then inspected for animals with abnormal appearance (e.g., injuries, abnormal swimming behavior, etc.). Overtly unhealthy looking tadpoles should be removed from the treatment tanks and replaced with larvae newly selected from the pooling tank.

Observations

26. For more in-depth information on test termination procedures and processing of tadpoles, refer to the OECD Guidance Document on Amphibian Thyroid Histology (9).

Day 7 Measurements

27. On day 7, five randomly chosen tadpoles per replicate are removed from each test tank. The random procedure used should give each organism on test equal probability of being selected. This can be achieved by using any randomizing method but requires that each tadpole be netted. Tadpoles not selected are returned to the tank of origin and the selected tadpoles are humanely euthanized in 150 to 200 mg/L e.g. MS-222, appropriately buffered with sodium bicarbonate to achieve pH 7.0. The euthanized tadpoles are rinsed in water and blotted dry, followed by body weight determination to the nearest milligram. Hind limb length, snout to vent length, and developmental stage (using a binocular dissection microscope) are determined for each tadpole.

Day 21 Measurements (Test Termination)

28. At test termination (day 21), the remaining tadpoles are removed from the test tanks and humanely euthanized in 150 to 200 mg/L e.g. MS-222, appropriately buffered with sodium bicarbonate, as above. Tadpoles are rinsed in water and blotted dry, followed by body weight determination to the nearest milligram. Developmental stage, SVL, and hind limb lengths are measured for each tadpole.

29. All larvae are placed in Davidson's fixative for 48 to 72 hours either as whole body samples or as trimmed head tissue samples containing the lower jaw for histological assessments. For histopathology, a total of five tadpoles should be sampled from each replicate tank. Since follicular cell height is stage dependent (10), the most appropriate sampling approach for histological analyses is to use stage-matched individuals, whenever possible. In order to select stage-matched individuals, all larvae should first be staged prior to selection and subsequent processing for data collection and preservation. This is necessary because normal divergence in development will result in differential stage distributions within each replicate tank.

30. Animals selected for histopathology (n=5 from each replicate) should be matched to the median stage of the controls (pooled replicates) whenever possible. If there are replicate tanks with more than five larvae at the appropriate stage, then five larvae are randomly selected.

31. If there are replicate tanks with less than five larvae at the appropriate stage, then randomly selected individuals from the next lower or upper developmental stage should be sampled to reach a total sample size of five larvae per replicate. Preferably, the decision to sample additional larvae from either the next lower or upper developmental stage should be made based on an overall evaluation of the stage distribution in the control and chemical treatments. That is, if the chemical treatment is associated with a retardation of development, than additional larvae should be sampled from the next lower stage. In turn, if the chemical treatment is associated with an acceleration of development, than additional larvae should be sampled from the next upper stage.

32. In cases of severe alterations of tadpole development due to treatment with a test chemical, there might be no overlap of the stage distribution in the chemical treatments with the calculated control median developmental stage. In only these cases, the selection process should be modified by using a stage different from the control median stage to achieve a stage-matched sampling of larvae for thyroid histopathology. Furthermore, if stages are indeterminate (i.e., asynchrony), then 5 tadpoles from each replicate should be randomly chosen for histological analysis. The rationale underlying sampling of any larvae that are not at a stage equivalent to the control median developmental stage should be reported.

Determination of Biological Endpoints

33. During the 21 day exposure phase, measurement of primary endpoints is performed on days 7 and 21, however daily observation of test animals is necessary. Table **3** provides an overview of the measurement endpoints and the corresponding observation time points. More detailed information for technical procedures for measurement of apical endpoints and histological assessments is available in the OECD guidance documents (9).

Table 3. Observation time points for primary endpoints in the AMA.

Apical Endpoints	Daily	Day 7	Day 21
-Mortality	●		
-Developmental Stage		●	●
-Hind Limb Length		●	●
-Snout-Vent Length		●	●
-Wet Body Weight		●	●
-Thyroid Gland Histology			●

Apical Endpoints

34. Developmental stage, hind limb length, SVL and wet weight are the apical endpoints of the AMA, and each is briefly discussed below. Further technical information for collecting these data is available in the guidance documents attached including procedures for computer-assisted analysis which are recommended for use.

Developmental Stage

35. The developmental stage of *X. laevis* tadpoles is determined using the staging criteria of Nieuwkoop and Faber (8). Developmental stage data are used to determine if development is accelerated, asynchronous, delayed or unaffected. Acceleration or delay of development is determined by making a comparison between the median stage achieved by the control and treated groups. Asynchronous development is reported when the tissues examined are not malformed or abnormal, but the relative timing of the morphogenesis or development of different tissues is disrupted within a single tadpole.

Hind Limb Length

36. Differentiation and growth of the hind limbs are under control of thyroid hormones and are major developmental landmarks already used in the determination of developmental stage. Hind limb development is used qualitatively in the determination of developmental stage, but is considered here as a quantitative endpoint. Therefore, hind limb length is measured as an endpoint to detect effects on the thyroid axis (Figure 2). For consistency, hind limb length is measured on the left hind limb. Hind limb length is evaluated both at day 7 and at day 21 of the test. On day 7, measuring hind limb length is straightforward, as illustrated in Figure 2. However, measuring hind limb length on day 21 is more complicated due to bends in the limb. Therefore, measurements of hind limb length at day 21 should originate at the body wall and follow the midline of the limb through any angular deviations. Changes in hind limb length at day 7, even if not evident at day 21, are still considered significant for potential thyroid activity. Length measurements are acquired from digital photographs using image analysis software as described in the OECD Guidance Document on Amphibian Thyroid Histopathology (9).

Body Length and Wet Weight

37. Determinations of snout to vent length (SVL) (Figure 2) and wet weight are included in the test protocol to assess possible effects of test substances on the growth rate of tadpoles in comparison to the control group and are useful in detecting generalized toxicity to the test compound. Because the removal of adherent water for weight determinations can cause stressful conditions for tadpoles and may cause skin damage, these measurements are performed on the day 7 sub-sampled tadpoles and all remaining tadpoles

at test termination (day 21). For consistency, use the cranial aspect of the vent as the caudal limit of the measurement.

38. Snout to vent length (SVL) is used to assess tadpole growth as illustrated in Figure 2,

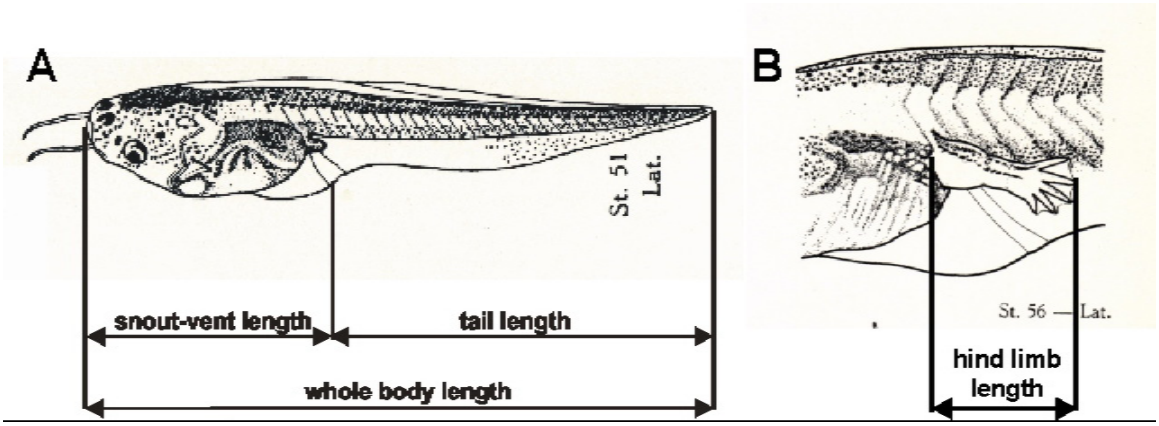


Figure 2. (A) Types of body length measurements and (B) Hind limb length measurements for *X. laevis* tadpoles (1).

Thyroid Gland Histology

39. While developmental stage and hind limb length are important endpoints to evaluate exposure-related changes in metamorphic development, developmental delay cannot, by itself, be considered a diagnostic indicator of anti-thyroidal activity. Some changes may only be observable by routine histopathological analysis. Diagnostic criteria include thyroid gland hypertrophy/atrophy, follicular cell hypertrophy, follicular cell hyperplasia, and as additional qualitative criteria: follicular lumen area, colloid quality and follicular cell height/shape. Severity grading (4 grades) should be reported. Information on obtaining and processing samples for histological analysis and for performing histologic analyses on tissue samples is available in "Amphibian Metamorphosis Assay: Part 1 - Technical guidance for morphologic sampling and histological preparation" and "Amphibian Metamorphosis Assay: Part 2 – Approach to reading studies, diagnostic criteria, severity grading and atlas"(9). Laboratories performing the assay for the first time(s) should seek advice from experienced pathologists for training purpose prior to undertaking histological analysis and evaluation of the thyroid gland. Overt and significant changes in apical endpoints indicating developmental acceleration or asynchrony may preclude the necessity to perform histopathological analysis of the thyroid glands. However, absence of overt morphological changes or evidence of developmental delay warrants histological analyses.

Mortality

40. All test tanks should be checked daily for dead tadpoles and the numbers recorded for each tank. The date, concentration and tank number for any observation of mortality should be recorded. Dead animals should be removed from the test tank as soon as observed. Mortality rates exceeding 10% may indicate inappropriate test conditions or toxic effects of the test chemical.

Additional Observations

41. Cases of abnormal behavior and grossly visible malformations and lesions should be recorded. The date, concentration and tank number for any observation of abnormal behavior, gross malformations, or lesions should be recorded. Normal behavior is characterized by the tadpoles being suspended in the water column with tail elevated above the head, regular rhythmic tail fin beating, periodic surfacing,

operculating, and being responsive to stimulus. Abnormal behavior would include, for example, floating on the surface, lying on the bottom of the tank, inverted or irregular swimming, lack of surfacing activity, and being nonresponsive to stimulus. In addition, gross differences in food consumption between treatments should be recorded. Gross malformations and lesions could include morphological abnormalities (e.g. limb deformities), hemorrhagic lesions, bacterial or fungal infections, to name a few. These determinations are qualitative and should be considered akin to clinical signs of disease/stress and made in comparison to control animals. If the occurrence or rate of occurrence is greater in exposed tanks than in the controls, then these should be considered as evidence for overt toxicity.

DATA AND REPORTING

Data Collection

42. All data should be collected using electronic or manual systems which conform to good laboratory practices (GLP). Study data should include:

Test substance:

- Characterization of the test substance: physical-chemical properties; information on stability and biodegradability;
- Chemical information and data: method and frequency of preparation of dilutions. Test chemical information includes actual and nominal concentrations of the test chemical, and in some cases, non-parent chemical, as appropriate. Test chemical measurements may be required for stock solutions as well as for test solutions;
- Solvent (if other than water): justification of the choice of solvent, and characterization of solvent (nature, concentration used);

Test conditions:

- Operational records: these consist of observations pertaining to the functioning of the test system and the supporting environment and infrastructure. Typical records include: ambient temperature, test temperature, photoperiod, status of critical components of the exposure system (e.g. pumps, cycle counters, pressures), flow rates, water levels, stock bottle changes, and feeding records. General water quality parameters include: pH, DO, conductivity, total iodine, alkalinity, and hardness;
- Deviations from the test method: this information should include any information or narrative descriptions of deviations from the test method;

Results:

- Biological observations and data: these include daily observations of mortality, food consumption, abnormal swimming behavior, lethargy, loss of equilibrium, malformations, lesions, etc. Observations and data collected at predetermined intervals include: developmental stage, hind limb length, snout vent length, and wet weight;
- Statistical analytical techniques and justification of techniques used; results of the statistical analysis preferably in tabular form;
- Histological data: these include narrative descriptions, as well as graded severity and incidence scores of specific observations, as detailed in the histopathology guidance document;
- Ad hoc observations: these observations should include narrative descriptions of the study that do not fit into the previously described categories.

Data reporting

43. Annex 2 contains daily data collection spreadsheets that can be used as guidance for raw data entry and for calculations of summary statistics. Additionally, reporting tables are provided that are convenient for communicating summaries of endpoint data. Reporting tables for histological assessments can be found in Annex 2.

Performance Criteria and Test Acceptability/Validity

44. Generally, gross deviations from the test method will result in unacceptable data for interpretation or reporting. Therefore, the following criteria in Table 4 have been developed as guidance for determining the quality of the test performed, the general performance of the control organisms.

Table 4. Performance criteria for the AMA.

Criterion	Acceptable limits
Test concentrations	Maintained at ≤ 20% CV(variability of measured test concentration) over the 21 day test
Mortality in controls	≤10% - mortality in any one replicate in the controls should not exceed 2 tadpoles
Minimum median developmental stage of controls at end of test	57
Spread of development stage in control group	The 10th and the 90th percentile of the development stage distribution should not differ by more than 4 stages
Dissolved Oxygen	≥ 40% air saturation*
pH	pH should be maintained between 6.5-8.5. The inter-replicate/inter-treatment differentials should not exceed 0.5.
Water temperature	22º ± 1ºC - the inter-replicate/inter-treatment differentials should not exceed 0.5 ºC
Test concentrations without overt toxicity	≥ 2
Replicate performance	≤ 2 replicates across the test can be compromised
Special conditions for use of a solvent	If a carrier solvent is used, both a solvent control and clean water control should be used and results reported
	Statistically significant differences between solvent control and water control groups are treated specially. See below for more information
Special conditions for static renewal system	Representative chemical analyses before and after renewal should be reported
	Ammonia levels should be measured immediately prior to renewal
	All water quality parameters listed in Table 1 of Annex 1 should be measured immediately prior to renewal
	Renewal period should not exceed 72 hours
	Appropriate feeding schedule (50% of the daily food ration of Sera Micron ®)

*Aeration of water can be maintained through bubblers. It is recommended to set bubblers at levels that do not create undue stress on the tadpoles.

Test Validity

45. The following requirements should be met to deem a test acceptable/valid:

Valid experiment in a test determined to be negative for thyroid activity:

1. For any given treatment (including controls), mortality cannot exceed 10%. For any given replicate, mortality cannot exceed three tadpoles, otherwise the replicate is considered compromised
2. At least two treatment levels, with all four uncompromised replicates, should be available for analysis
3. At least two treatment levels without overt toxicity should be available for analysis

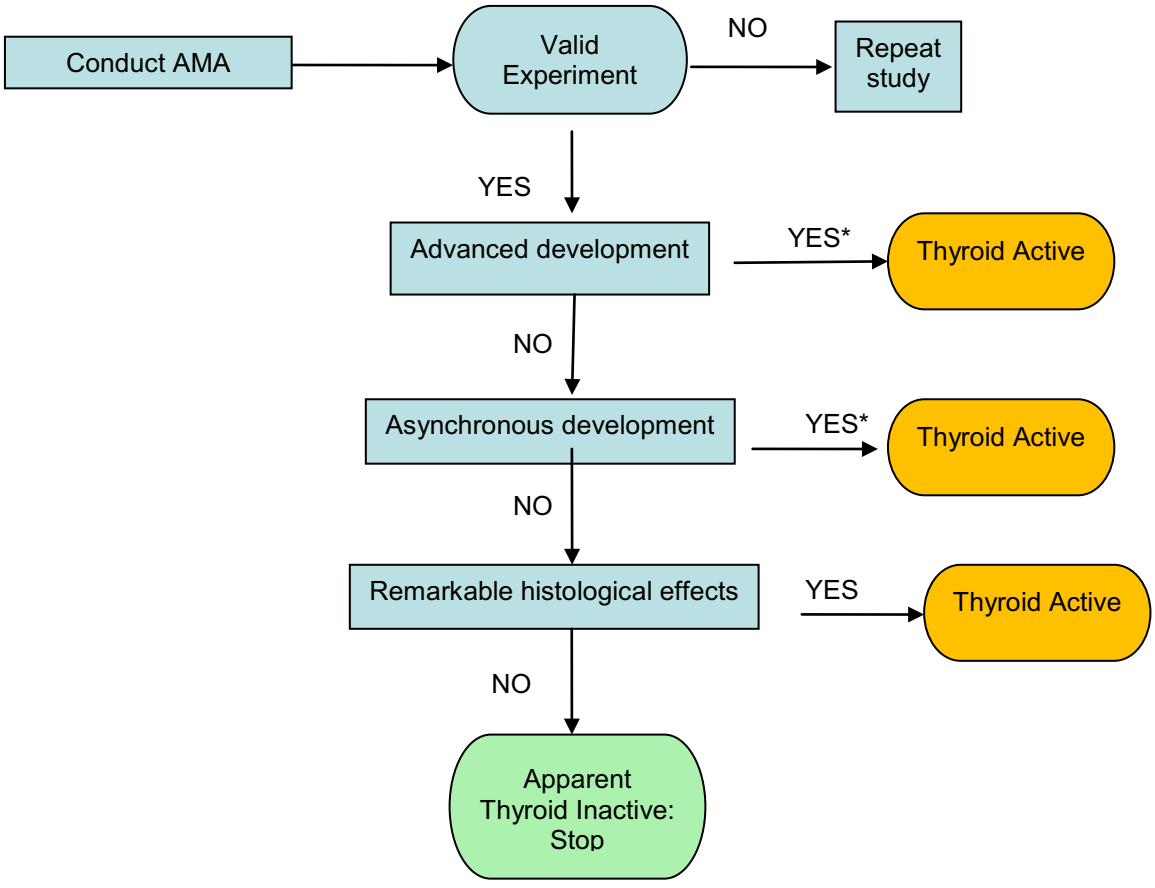
Valid experiment in a test determined to be positive for thyroid activity:

1. Mortality of no more than two tadpoles/replicate in the control group can occur

Decision logic for the conduct of the AMA

46. Decision logic was developed for the AMA to provide logical assistance in the conduct and interpretation of the results of the bioassay (see flow chart in Figure 3). The decision logic, in essence, weighs the endpoints in that advanced development, asynchronous development, and thyroid histopathology are weighed heavily, while delayed development, snout-vent length and wet body weight, parameters that can potentially be affected by general toxicity, are weighed less heavily.

Figure 3. Decision logic for the conduct of the AMA.



*Histology may be required by some regulatory authorities despite significant differences in advanced and asynchronous development. The entity performing this test is encouraged to consult the necessary authorities prior to the performing the test to determine which endpoints are required.

Advanced development (determined using developmental stage, SVL and HLL)

47. Advanced development is only known to occur through effects which are thyroid hormone related. These may be peripheral tissue effects such as direct interaction with the thyroid hormone receptor (such as with T4) or effects which alter circulating thyroid hormone levels. In either case, this is considered sufficient evidence to indicate that the chemical has thyroid activity. Advanced development is evaluated in one of two ways. First, the general developmental stage can be evaluated using the standardized approach detailed in Nieuwkoop and Faber (8). Second, specific morphological features may be quantified, such as hind limb length, at both days 7 and 21, which is positively associated with agonistic effects on the thyroid hormone receptor. If statistically significant advances in development or hind limb length occur, then the test indicates that the chemical is thyroid active.

48. The evaluation of test animals for the presence of accelerated development relative to the control population will be based on results of statistical analyses performed for the following four endpoints:

- hind limb length (normalized by SVL) on study day 7
- hind limb length (normalized by SVL) on study day 21
- developmental stage on study day 7
- developmental stage on study day 21.

49. Statistical analyses of hind limb length should be performed based on measurements of the length of the left hind limb. Hind limb length is normalized by taking the ratio hind limb length to snout-to-vent length of an individual. The mean of the normalized values for each treatment level are then compared. Acceleration of development is then indicated by a significant increase of mean hind limb length (normalized) in a chemical treatment group compared to the control group on study day 7 and/or study day 21 (see Annex 3).

50. Statistical analyses of developmental stage should be performed based on determination of developmental stages according to the morphological criteria described by Nieuwkoop and Faber (7). Acceleration of development is indicated when the multi-quantal analysis detects a significant increase of developmental stage values in a chemical treatment group compared to the control group on study day 7 and/or study day 21.

51. In the AMA test method, a significant effect on any of the four endpoints mentioned above is regarded sufficient for a positive detection of accelerated development. That is, significant effects on hind limb length at a specific time point do not require corroboration by significant effects on hind limb length at the alternative time point nor by significant effects on developmental stage at this specific time point. In turn, significant effects on developmental stage at a specific time point do not require corroboration by significant effects at developmental stage on the alternative time point nor by significant effects on hind limb length at this specific time point. The weight of evidence for accelerated development will nevertheless increase if significant effects are detected for more than one endpoint.

Asynchronous development (determined using developmental stage criteria)

52. Asynchronous development is characterized by disruption of the relative timing of the morphogenesis or development of different tissues within a single tadpole. The inability to clearly establish the developmental stage of an organism using the suite of morphological endpoints considered typical of any given stage indicates that the tissues are developing asynchronously through metamorphosis. Asynchronous development is an indicator of thyroid activity. The only known modes of action causing

asynchronous development are through effects of chemicals on peripheral thyroid hormone action and/or thyroid hormone metabolism in developing tissues such as is observed with deiodinase inhibitors.

53. The evaluation of test animals for the presence of asynchronous development relative to the control population will be based on gross morphological assessment of test animals on study day 7 and study day 21.

54. The description of normal development of *Xenopus laevis* by Nieuwkoop and Faber (8) provides the framework for identifying a sequential order of normal tissue remodelling. The term "asynchronous development" refers specifically to those deviations in tadpole gross morphological development that disallow the definitive determination of a developmental stage according to the criteria of Nieuwkoop and Faber (8) because key morphological landmarks show characteristics of different stages.

55. As implicated by the term "asynchronous development", only cases showing deviations in the progress of remodelling of specific tissues relative to the progress of remodelling of other tissues should be considered. Some classical phenotypes include delay or absence of fore limb emergence despite normal or advanced development of hind limbs and tail tissues, or the precocious resorption of gills relative to the stage of hind limb morphogenesis and tail resorption. An animal will be recorded as showing asynchronous development if it cannot be assigned to a stage because it fails to meet a majority of of the landmark developmental criteria for a given Niewkoop and Faber stage (8), or if there is extreme delay or acceleration of one or more key features (e.g. tail completely resorbed, but forelimbs not emerged). This assessment is performed qualitatively and should examine the full suite of landmark features listed by Nieuwkoop and Faber (8). However it is not necessary to record the developmental state of the various landmark features of animals being observed. Animals recorded as showing asynchronous development are not assigned to a Nieuwkoop and Faber (8) development stage.

56. Thus, a central criterion for designating cases of abnormal morphological development as "asynchronous development" is that the relative timing of tissue remodelling and tissue morphogenesis is disrupted whereas the morphology of affected tissues is not overtly abnormal. One example to illustrate this interpretation of gross morphological abnormalities is that retarded hind limb morphogenesis relative to development of other tissues will fulfill the criterion of "asynchronous development" whereas cases showing missing hind limbs, abnormal digits (e.g. ectrodactyly, polydactyly), or other overt limb malformations should not be considered as "asynchronous development".

57. In this context, the major morphological landmarks that should be evaluated for their coordinated metamorphic progress should include hind limb morphogenesis, fore limb morphogenesis, fore limb emergence, the stage of tail resorption (particularly the resorption of the tail fin), and head morphology (e.g. gill size and stage of gill resorption, lower jaw morphology, protrusion of Meckel's cartilage).

58. Dependent on the mode of chemical action, different gross morphological phenotypes can occur. Some classical phenotypes include delay or absence of fore limb emergence in spite of normal or advanced development of hind limbs and tail tissues, precocious gill resorption relative to hind limb and tail remodelling.

Histopathology

59. If the chemical does not cause overt toxicity and does not accelerate development or cause asynchronous development, then histopathology of the thyroid glands is evaluated using the appropriate guidance document (9). Developmental retardation, in the absence of toxicity, is a strong indicator of anti-thyroid activity, but the developmental stage analysis is less sensitive and less diagnostic than the histopathological analysis of the thyroid gland. Therefore, conducting histopathological analyses of the

thyroid glands is required in this case. Effects on thyroid gland histology have been demonstrated in the absence of developmental effects. If changes in thyroid histopathology occur, then the chemical is considered to be thyroid active. If no developmental delays or histological lesions are observed in the thyroid glands, then the chemical is considered to be thyroid inactive. The rationale for this decision is that the thyroid gland is under the influence of TSH and any chemical which alters circulating thyroid hormone sufficiently to alter TSH secretion will result in histopathological changes in the thyroid glands. Various modes and mechanisms of action can alter circulating thyroid hormone. So, while thyroid hormone level is indicative of a thyroid related effect, it is insufficient to determine which mode or mechanism of action is related to the response.

60. Because this endpoint is not amenable to basic statistical approaches, the determination of an effect associated with exposure to a chemical shall be made through expert opinion by a pathologist.

Delayed development (determined using developmental stage, HLL, BW, SVL)

61. Delayed development can occur through anti-thyroidal mechanisms and through indirect toxicity. Mild developmental delays coupled with overt signs of toxicity likely indicate a non-specific toxic effect. Evaluation of non-thyroidal toxicity is an essential element of the test to reduce the probability of false positive outcomes. Excessive mortality is an obvious indication that other toxic mechanisms are occurring. Similarly, mild reductions in growth, as determined by wet weight and/or SVL length, also suggest non-thyroidal toxicity. Apparent increases in growth are commonly observed with compounds that negatively affect normal development. Consequently, the presence of larger animals does not necessarily indicate non-thyroidal toxicity. However, growth should never be solely relied upon to determine thyroid toxicity. Rather, growth, in conjunction with developmental stage and thyroid histopathology, should be used to determine thyroid activity. Other endpoints should also be considered in determining overt toxicity including edema, hemorrhagic lesions, lethargy, reduced food consumption, erratic/altered swimming behavior, etc. If all test concentrations exhibit signs of overt toxicity, the test compound should be re-evaluated at lower test concentrations before determining whether the compound is potentially thyroid active or thyroid inactive.

62. Statistically significant developmental delays, in absence of other signs of overt toxicity, indicate that the chemical is thyroid active (antagonistic). In the absence of strong statistical responses, this outcome may be augmented with results from thyroid histopathology.

Statistical analyses

63. Statistical analyses of the data should preferably follow procedures described in the document Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application (11). For all continuous quantitative endpoints (HHL, SVL, wet weight) consistent with a monotone dose-response, the Jonckheere-Terpstra test should be applied in step-down manner to establish a significant treatment effect.

64. For continuous endpoints that are not consistent with a monotone dose-response, the data should be assessed for normality (preferably using the Shapiro-Wilk or Anderson-Darling test) and variance homogeneity (preferably using the Levene test). Both tests are performed on the residuals from an ANOVA. Expert judgment can be used in lieu of these formal tests for normality and variance homogeneity, though formal tests are preferred. Where non-normality or variance heterogeneity is found, a normalizing, variance stabilizing transformation should be sought. If the data (perhaps after a transformation) are normally distributed with homogeneous variance, a significant treatment effect is determined from Dunnett's test. If the data (perhaps after a transformation) are normally distributed with heterogeneous variance, a significant treatment effect is determined from the Tamhane-Dunnett or T3 test or from the Mann-Whitney-Wilcoxon U test. Where no normalizing transformation can be found, a

significant treatment effect is determined from the Mann-Whitney-Wilcoxon U test using a Bonferroni-Holm adjustment to the p-values. The Dunnett test is applied independently of any ANOVA F-test and the Mann-Whitney test is applied independently of any overall Kruskall-Wallis test.

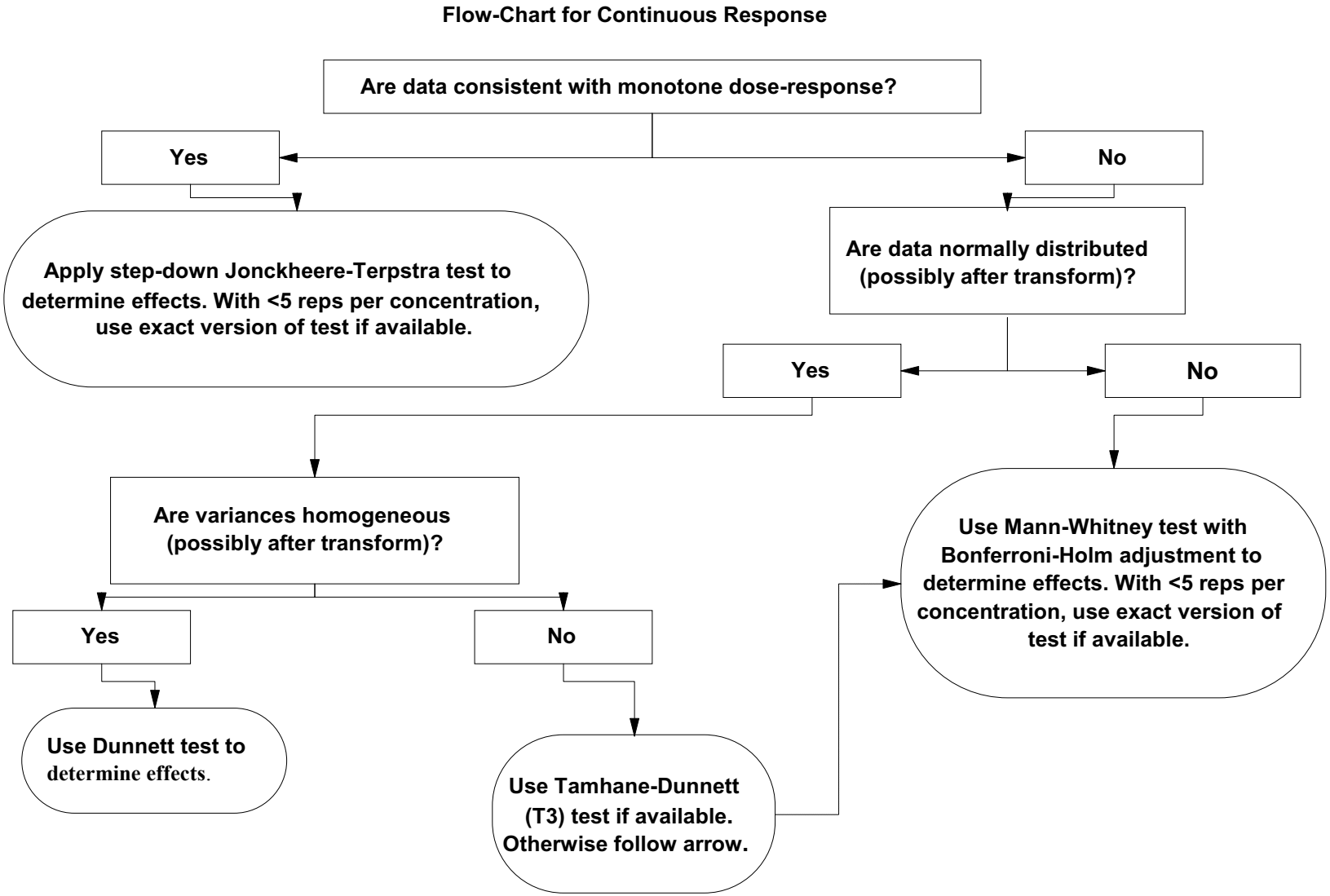
65. Significant mortality is not expected but should be assessed from the step-down Cochran-Armitage test where the data are consistent with dose-response monotonicity, and otherwise from Fisher's Exact test with a Bonferroni-Holm adjustment.

66. A significant treatment effect for developmental stage is determined from the step-down application of the Jonckheere-Terpstra test applied to the replicate medians. Alternatively, and preferably, the multi-quantal Jonckheere test from the 20^{th} to the 80^{th} percentile should be used for effect determination, as it takes into account changes to the distribution profile.

67. The appropriate unit of analysis is the replicate so the data consist of replicate medians if the Jonckheere-Terpstra or Mann-Whitney U test is used, or the replicate means if Dunnett's test is used. Dose-response monotonicity can be assessed visually from the replicate and treatment means or medians or from formal tests such as previously described (11). With fewer than five replicates per treatment or control, the exact permutation versions of the Jonckheere-Terpstra and Mann-Whitney tests should be used if available. The statistical significance of all tests indicated is judged at the 0.05 significance level.

68. Figure 4 is a flow chart for performing statistical tests on continuous data.

Figure 4. Flowchart for statistical approaches for continuous response data.



) OECD, (2009)

Special data analysis considerations

Use of compromised treatment levels

69. Several factors are considered when determining whether a replicate or entire treatment demonstrates overt toxicity and should be removed from analysis. Overt toxicity is defined as >2 mortalities in any replicate that can only be explained by toxicity rather than technical error. Other signs of overt toxicity include hemorrhage, abnormal behaviors, abnormal swimming patterns, anorexia, and any other clinical signs of disease. For sub-lethal signs of toxicity, qualitative evaluations may be necessary, and should always be made in reference to the clean water control group.

Solvent controls

70. The use of a solvent should only be considered as a last resort, when all other chemical delivery options have been considered. If a solvent is used, then a clean water control should be run in concert. At the termination of the test, an evaluation of the potential effects of the solvent should be performed. This is done through a statistical comparison of the solvent control group and the clean water control group. The most relevant endpoints for consideration in this analysis are developmental stage, SVL, and wet weight, as these can be affected through non-thyroidal toxicities. If statistically significant differences are detected in these endpoints between the clean water control and solvent control groups, determine the study endpoints for the response measures using the clean water control. If there is no statistically significant difference between the clean water control and solvent control for all measured response variables, determine the study endpoints for the response measures using the pooled dilution-water and solvent controls.

Treatment groups achieving developmental stage 60 and above

71. After stage 60, tadpoles show a reduction in size and weight due to tissue resorption and reduction of absolute water content. Thus, measurements of wet weight and SVL cannot appropriately be used in statistical analyses for differences in growth rates. Therefore, wet weight and length data from organisms >NF60 should be censored and cannot be used in analyses of replicate means or replicate medians. Two different approaches could be used to analyse these growth-related parameters.

72. One approach is to consider only tadpoles with developmental stages lower or equal to stage 60 for the statistical analyses of wet weight and/or SVL. This approach is believed to provide sufficiently robust information about the severity of possible growth effects as long as only a small proportion of test animals are removed from the analyses (≤20%). If an increased number of tadpoles show development beyond stage 60 (≥20%) in one or more nominal concentration(s), then a two -factor ANOVA with a nested variance structure should be undertaken on all tadpoles to assess growth effects due to chemical treatments while taking into account the effect of late stage development on growth. Annex 3 provides guidance on the two-factor ANOVA analysis of weight and length. .

LITERATURE

1. OECD (2004) Report of the Validation of the Amphibian Metamorphosis Assay for the detection of thyroid active substances: Phase 1 – Optimisation of the Test Protocol. Environmental Health and Safety Publications. Series on Testing and Assessment. No. 77, Paris.

2. OECD (2007) Final Report of the Validation of the Amphibian Metamorphosis Assay: Phase 2 - Multi-chemical Interlaboratory Study. Environmental Health and Safety Publications. Series on Testing and Assessment. No. 76. Paris

3. OECD (2008) Report of the Validation Peer Review for the Amphibian Metamorphosis Assay and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up of this Report. Environmental Health and Safety Publications. Series on Testing and Assessment. No. 92. Paris

4. OECD (2000) Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures. Environmental Health and Safety Publications. Series on Testing and Assessment. No. 23. Paris

5. ASTM (2002) Standard Guide for Conducting Acute Toxicity Tests on Test Materials with Fishes, Macroinvertebrates, and Amphibians. American Society for Testing and Materials, ASTM E729-96(2002), Philadelpia, PA

6. ASTM (2004) Standard Guide for Conducting the Frog Embryo Teratogenesis Assay - Xenopus (FETAX). E 1439-98

7. Kahl,M.D., Russom,C.L., DeFoe,D.L. & Hammermeister,D.E. (1999) Saturation units for use in aquatic bioassays. Chemosphere 39, pp. 539-551

8. Nieuwkoop,P.D. & Faber,J. (1994) Normal Table of Xenopus laevis. Garland Publishing, New York

9. OECD (2007) Guidance Document on Amphibian Thyroid Histology. Environmental Health and Safety Publications. Series on Testing and Assessment. No. 82. Paris

10. Dodd,M.H.I. & Dodd,J.M. (1976) Physiology of Amphibia. Lofts,B. (ed.), Academic Press, New York, pp. 467-599

11. OECD (2006) Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application. Environmental Health and Safety Publications. Series on Testing and Assessment, No. 54. Paris

12. Hutchinson TH, Shillabeer N, Winter MJ, Pickford DB, 2006. Acute and chronic effects of carrier solvents in aquatic organisms: A critical review. Review. Aquatic Toxicology, 76 ; pp.69–92.

ANNEX 1

Table 1: Experimental Conditions for the 21-day Amphibian Metamorphosis Assay

Test Animal		*Xenopus laevis* larvae	
Initial Larval Stage		Nieuwkoop and Faber stage 51	
Exposure Period		21 days	
Larvae Selection Criteria		Developmental stage and total length (optional)	
Test Concentrations		Minimum of 3 concentrations spanning approximately one order of magnitude	
Exposure Regime		Flow-through (preferred) and/or static-renewal	
Test System Flow-Rate		25 mL/min (complete volume replacement ca. every 2.7 h)	
Primary Endpoints / Determination Days		Mortality	Daily
		Developmental Stage	D 7 and 21
		Hind Limb Length	D 7 and 21
		Snout-Vent Length	D 7 and 21
		Wet Body Weight	D 7 and 21
		Thyroid Histology	D 21
Dilution Water / Laboratory Control		Dechlorinated tap water (charcoal-filtered) or the equivalent laboratory source	
Larval Density		20 larvae / test vessel (5 / L)	
Test Solution / Test Vessel		4-10 L (10-15 cm minimum water) / Glass or Stainless Steel test vessel (e.g., 22.5 cm x 14 cm x 16.5 cm)	
Replication		4 replicate test vessels / test concentration and control	
Acceptable Mortality Rate in Controls		≤ 10% per replicate test vessel	
Thyroid Fixation	Number Fixed	All tadpoles (5/replicate are evaluated initially)	
	Region	Head or whole body	
	Fixation Fluid	Davidson's fixative	
Feeding	Food	Sera Micron® or equivalent	
	Amount / Frequency	See Table 1 for feeding regime using Sera Micron®	
Lighting	Photoperiod	12 h Light : 12 h dark	
	Intensity	600 to 2000 lux (Measured at Water Surface)	
Water Temperature		22º ± 1ºC	
pH		6.5 – 8.5	
Dissolved Oxygen (DO) Concentration		>3.5 mg/L (>40% Air Saturation)	
Analytical Chemistry Sample Schedule		Once / Week (4 Sample Events / Test)	

ANNEX 2

Reporting tables for raw data and summary data

Annex 2- Table 1: General test chemical information

Chemical Information:

Enter Test Substance, Concentration Units, and Treatments

Test substance:	
Concentration units:	
Treatment 1:	
Treatment 2:	
Treatment 3:	
Treatment 4:	

Date (day 0):	00/00/00	Enter date (mm/dd/yy)
Date (day 7):	00/00/00	Enter date (mm/dd/yy)
Date (day 21):	00/00/00	Enter date (mm/dd/yy)

Annex 2- Table 2: Raw data collection sheets for days 7 and 21

DAY X

DATE 00/00/00

ROW	Concentration TRT	Treatment Number TRT #	Replicate Number REP	Individual Number IND	Individual Identifier ID #	Developmental Stage STAGE	SVL Length (mm) BL	Hindlimb Length (mm) HLL	Whole Organism Wet Weight (mg) WEIGHT
1	0.00	1							
2	0.00	1							
3	0.00	1							
4	0.00	1							
5	0.00	1							
6	0.00	1							
7	0.00	1							
8	0.00	1							
9	0.00	1							
10	0.00	1							
11	0.00	1							
12	0.00	1							
13	0.00	1							
14	0.00	1							
15	0.00	1							
16	0.00	1							
17	0.00	1							
18	0.00	1							
19	0.00	1							
20	0.00	1							
21	0.00	2							
22	0.00	2							
23	0.00	2							
24	0.00	2							
25	0.00	2							
26	0.00	2							
27	0.00	2							
28	0.00	2							
29	0.00	2							
30	0.00	2							
31	0.00	2							
32	0.00	2							
33	0.00	2							
34	0.00	2							
35	0.00	2							
36	0.00	2							
37	0.00	2							
38	0.00	2							
39	0.00	2							
40	0.00	2							
41	0.00	3							
42	0.00	3							
43	0.00	3							
44	0.00	3							
45	0.00	3							
46	0.00	3							
47	0.00	3							
48	0.00	3							
49	0.00	3							
50	0.00	3							
51	0.00	3							
52	0.00	3							
53	0.00	3							
54	0.00	3							
55	0.00	3							
56	0.00	3							
57	0.00	3							
58	0.00	3							
59	0.00	3							
60	0.00	3							
61	0.00	4							
62	0.00	4							
63	0.00	4							
64	0.00	4							
65	0.00	4							
66	0.00	4							
67	0.00	4							
68	0.00	4							
69	0.00	4							
70	0.00	4							
71	0.00	4							
72	0.00	4							
73	0.00	4							
74	0.00	4							
75	0.00	4							
76	0.00	4							
77	0.00	4							
78	0.00	4							
79	0.00	4							
80	0.00	4							

Annex 2- Table 3: Calculated summaries for endpoint data from days 7 and 21

		Developmental Stage			SVL (mm)		Hindlimb Length (mm)		Weight (mg)	
TRT	REP	MIN	MEDIAN	MAX	MEAN	STD DEV	MEAN	STD DEV	MEAN	STD DEV
1	1	0	#NUM!	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
1	2	0	#NUM!	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
1	3	0	#NUM!	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
1	4	0	#NUM!	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
2	1	0	#NUM!	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
2	2	0	#NUM!	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
2	3	0	#NUM!	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
2	4	0	#NUM!	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
3	1	0	#NUM!	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
3	2	0	#NUM!	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
3	3	0	#NUM!	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
3	4	0	#NUM!	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
4	1	0	#NUM!	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
4	2	0	#NUM!	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
4	3	0	#NUM!	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
4	4	0	#NUM!	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!

Note: Cell calculations are associated with data entries into Table 2.

Annex 2- Table 4: Daily mortality data

Test Day	Date	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
0	00/00/00																
1	#VALUE!																
2	#VALUE!																
3	#VALUE!																
4	#VALUE!																
5	#VALUE!																
6	#VALUE!																
7	#VALUE!																
8	#VALUE!																
9	#VALUE!																
10	#VALUE!																
11	#VALUE!																
12	#VALUE!																
13	#VALUE!																
14	#VALUE!																
15	#VALUE!																
16	#VALUE!																
17	#VALUE!																
18	#VALUE!																
19	#VALUE!																
20	#VALUE!																
21	#VALUE!																
Replicate count		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Treatment count		0				0				0				0			

Note: Cell calculations are associated with data entries into Table 1.

Annex 2 – Table 5: Water Quality Criteria

Exposure System (flow-through/static renewal):
Temperature:
Light intensity:
Light-dark cycle:
Food:
Feeding rate:
water pH:
Iodine concentration in test water:

Annex 2- Table 6: Summary chemistry data

Chemical name:																				
CAS #:																				
Test Day	Date	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
0	00/00/00																				
1	#VALUE!																				
2	#VALUE!																				
3	#VALUE!																				
4	#VALUE!																				
5	#VALUE!																				
6	#VALUE!																				
7	#VALUE!																				
8	#VALUE!																				
9	#VALUE!																				
10	#VALUE!																				
11	#VALUE!																				
12	#VALUE!																				
13	#VALUE!																				
14	#VALUE!																				
15	#VALUE!																				
16	#VALUE!																				
17	#VALUE!																				
18	#VALUE!																				
19	#VALUE!																				
20	#VALUE!																				
21	#VALUE!																				

Note: Cell calculations are associated with data entries into Table 1.

Annex 2- Table 7: Histopathology reporting tables for core criteria

Date: Chemical: Pathologist:

		Thyroid gland hypertrophy	Thyroid gland atrophy	Follicular cell hypertrophy	Follicular cell hyperplasia
Control Animal ID - replicate 1					
Control Animal ID - replicate 2					
Total:					

		Thyroid gland hypertrophy	Thyroid gland atrophy	Follicular cell hypertrophy	Follicular cell hyperplasia
___ Dose Animal ID - replicate 1					
___ Dose Animal ID - replicate 2					
Total:					

		Thyroid gland hypertrophy	Thyroid gland atrophy	Follicular cell hypertrophy	Follicular cell hyperplasia
___ Dose Animal ID - replicate 1					
___ Dose Animal ID - replicate 2					
Total:					

		Thyroid gland hypertrophy	Thyroid gland atrophy	Follicular cell hypertrophy	Follicular cell hyperplasia
___ Dose Animal ID - replicate 1					
___ Dose Animal ID - replicate 2					
Total:					

Annex 2-Table 8: Additional histopathology criteria

Date: Chemical: Pathologist:

		Follicular lumen area increase	Follicular lumen area decrease
Control Animal ID - replicate 1			
Control Animal ID - replicate 2			
Total:			

		Follicular lumen area increase	Follicular lumen area decrease
____ Dose Animal ID - replicate 1			
____ Dose Animal ID - replicate 2			
Total:			

		Follicular lumen area increase	Follicular lumen area decrease
____ Dose Animal ID - replicate 1			
____ Dose Animal ID - replicate 2			
Total:			

		Follicular lumen area increase	Follicular lumen area decrease
____ Dose Animal ID - replicate 1			
____ Dose Animal ID - replicate 2			
Total:			

Annex 2-Table 9: Narrative descriptions for histopathological findings

Date:
Chemical:
Pathologist:

		Narrative description
Control Animal ID - replicate 1		
Control Animal ID - replicate 2		

____ Dose ____ Animal ID - replicate 1		
____ Dose ____ Animal ID - replicate 2		

____ Dose ____ Animal ID - replicate 1		
____ Dose ____ Animal ID - replicate 2		

____Dose ____Animal ID - replicate 1		
____Dose ____Animal ID - replicate 2		

Annex 2-Table 10: Summary reporting table template for day x (7 or 21) of the AMA

		Control				Dose 1					Dose 2					Dose 3				
Endpoint	**Replicate**	**Mean**	**SD**	**CV**	**N**	**Mean**	**SD**	**CV**	**N**	***p*-value**	**Mean**	**SD**	**CV**	**N**	***p*-value**	**Mean**	**SD**	**CV**	**N**	***p*-value**
Hind Limb Length (mm)	1																			
	2																			
	3																			
	4																			
	Mean:																			
SVL (mm)	1																			
	2																			
	3																			
	4																			
	Mean:																			
Wet Weight (mg)	1																			
	2																			
	3																			
	4																			
	Mean:																			

Annex 2- Table 11: Summary reporting table template for day x (7 or 21) developmental stage data for the AMA

		Control				**Dose 1**					**Dose 2**					**Dose 3**				
	Replicate	**Median**	**Min**	**Max**	**N**	**Median**	**Min**	**Max**	**N**	***p*-value**	**Median**	**Min**	**Max**	**N**	***p*-value**	**Median**	**Min**	**Max**	**N**	***p*-value**
Developmental Stage	1																			
	2																			
	3																			
	4																			
	Mean:																			

ANNEX 3

Alternative Analysis of weight and length in the case of late stage development exceeding 20% of tadpoles in one or more concentration(s)

If an increased number of tadpoles show development beyond stage 60 (≥20%) in one or more nominal concentration(s), then a two-factor ANOVA with a nested variance structure should be undertaken on all tadpoles to assess growth effects due to chemical treatments while taking into account the effect of late stage development on growth.

The proposal is to use all data but take into account the effect of late stage development. This can be done with a two-factor ANOVA with a nested variance structure. Define LateStage='Yes' for an animal if its developmental stage is 61 or greater. Otherwise, define LateStage='No'. Then a two-factor ANOVA with concentration and LateStage and their interaction can be done, with Rep(Conc) a random factor and Tadpole(Rep) another random effect. This still treats the rep as the unit of analysis and gives essentially the same results as a weighted analysis of rep*latestage means, weighted by the number of animals per mean. If the data violate the normality or variance homogeneity requirements of ANOVA, then a normalized rank-order transform can be done to remove that objection.

In addition to the standard ANOVA F-tests for the effects of Conc, LateStage, and their interactions, the interaction F-test can be "sliced" into two additional ANOVA F-test can be done, one on the mean responses across concentrations for LateStage='No' and another on the mean responses across concentrations for LateStage='Yes.' Further comparisons of treatment means against control are done within each level of LateStage. A trend-type analysis can be done using appropriate contrasts or simple pairwise comparisons can be done if there is evidence of non-monotone dose-response within a level of the LateStage variable. A Bonferroni-Holm adjustments to the p-values is made only if the corresponding F-slice is not significant. This can be done in SAS and, presumably, other statistical software packages. Complications can arise when there are no late stage animals in some concentrations, but these situations can be handled in a straight-forward fashion.