

## **OECD GUIDELINES FOR THE TESTING OF CHEMICALS**

### **PROPOSAL FOR A NEW GUIDELINE 223**

#### **Avian Acute Oral Toxicity Test**

#### **INTRODUCTION**

1. This Test Guideline is designed to estimate the acute oral toxicity of substances to birds. This Guideline provides a sequential testing procedure that optimises the placement of doses and matches the precision of the endpoint with the precision required for hazard assessment and labelling. The method has been designed in a way to minimise the numbers of birds used.

2. This Guideline started development at the SETAC/OECD Workgroup on avian toxicity testing following a workshop held in Pensacola, United States, in 1994 (1) with subsequent SETAC meetings in Europe and the United States to develop and optimise the sequential testing design. The optimal sequential testing design has been developed with extensive statistical validation (2).

#### **INITIAL CONSIDERATIONS**

3. The information required by different hazard assessment schemes may vary considerably. To satisfy these various needs, three alternative tests are described:

- **Limit dose test** – this is the preferred test when toxicity is expected to be low and lethality is unlikely at the limit dose. The limit dose must be adequate for assessment purposes, and it is usually 2000 mg/kg body-weight.
- **Full test: LD50-only test** – this is the preferred test when an estimate of the median lethal dose is required but neither the slope of the dose response curve or the confidence interval for the LD50 is required. This may be the appropriate test to estimate a percentile of a species sensitivity distribution of LD50s and to provide information for product labelling purposes.
- **Full test: Dose-response test** – this is the preferred test when the slope of the dose response curve and/or the confidence interval is required in addition to an estimate of the LD50.

4. Definitions used in this Guideline are given in Annex 1.

#### **PRINCIPLE OF THE TEST**

5. The test is divided into a number of discrete stages. At each stage a number of birds are simultaneously given a single dose (mg/kg bodyweight) of the test substance into the crop or proventriculus. At each stage, individual birds may receive different doses or doses may be replicated. The recommended strategy for testing materials that are unlikely to present a significant hazard is to perform a test with multiple birds dosed at the limit dose. If toxicity is expected the recommended strategy is to use non-replicated doses in the first two stages and to use replicates of only two doses at the third and later stages. In the first stage, the range of doses is based on the best available estimate of the LD50 – for example, the rodent LD50. Doses for subsequent stages are based on the mortalities observed in all previous stages.

## **PROPOSAL B**

6. Since mortality is the primary endpoint and background mortality is presumed to be negligible, no controls are used.

7. After dosing, the birds are observed for a 14 day period in order to measure mortality. It may be necessary to extend the observation period depending on evidence of delayed effects. The staged test design is easiest to apply to chemicals that produce death within a few days. In such cases it is not necessary to wait 14 days before moving to the next stage. Mortality observed after only three days may be used to determine doses for the following stage.

8. During the test, animals obviously in pain or showing signs of severe distress should be humanely killed.

### **DESCRIPTION OF THE TEST METHODS**

#### **Selection of bird species**

9. The preferred bird species are bobwhite quail, *Colinus virginianus* and Japanese quail, *Coturnix coturnix japonica* (*Galliform*). Mallard, *Anas platyrhynchos* (*Anseriform*) is not a preferred species because it is prone to regurgitate the dose, and only recommended if a second species is necessary.

10. Occasionally, it may be necessary to test additional species in order to take account of the species sensitivity. When possible captive bred species should be used. If this is not possible, species that are easily caught, abundant and acclimatise well to test conditions may be used. A short list of species that may be considered, in addition to quail, include mallard, pigeon, *Columba livia* (*Collumbiform*), zebra finch, *Poephila guttata* (*Passeriform*) and budgerigar, *Melopsittacus undulatus* (*Psittaciform*).

11. Birds should be in mature plumage but not in breeding condition. Wild phenotypes are preferred, where possible. Captive bred birds should be from the same source and breeding population, and when possible, breeding history should demonstrate periodic out-breeding to maintain genetic heterogeneity.

12. Birds should be drawn at random from a group comprising a single sex. Either sex could be used. If it is not possible to separate sexes by plumage, birds can be drawn at random from the whole population. If it is necessary to check for sensitivity due to sex an additional stage could be added using 4-6 birds of the opposite sex covering the critical dose range (2-3 sequential doses either side of the estimated mean).

#### **Housing and test conditions**

13. Individual caging is preferred to identify animals regurgitating the dose and to prevent fighting. However, group caging may be used if it improves animal welfare in the case of sociable species. Conditions should be within optimal limits for the test species. Cages must be constructed from inert material that can withstand regular cleaning after each test. They must be cleaned thoroughly and rinsed or steam cleaned before use. Birds should be kept in cages or pens of suitable size for the species being tested. Minimum floor areas recommended are 3000cm<sup>2</sup> for pigeon; 2000cm<sup>2</sup> for mallard; 1000cm<sup>2</sup> for quail and 500cm<sup>2</sup> for budgerigar and zebra finch. Cage floors should be constructed of mesh, large enough to allow faeces to fall through, but not restrictive on the bird's movements. Pigeons, zebra finches and budgerigars require perches

14. The test environment may be under controlled conditions or at ambient temperature and humidity. Temperatures within the range 15-27<sup>0</sup>C are suitable for quail and duck but should fluctuate

## **PROPOSAL B**

as little as possible during the test stages. Ventilation should be sufficient to supply at least 10 changes of air/hour. The photoperiod for quail and mallard should be 8 hours light and 16 hours dark. For other species it may be necessary to increase the light phase to 10 hours. Fresh food and water should be provided *ad libitum*. Commercial gamebird diets can be used, but they must be nutritionally appropriate for the species involved. Medication should be avoided within 14 days of dosing or during the observation period. Diets and water should be periodically analysed to check for impurities.

### **Preparation of birds**

15. Birds should be uniquely identified. Acclimatisation to test conditions and diet prior to dosing should be at least 14 days for cage reared and longer for wild caught birds. All birds must be in healthy condition and should not be used if greater than 5% of cage-reared and greater than 10% of wild test birds die during the acclimatisation period. If wild birds do not acclimatise they should be released. Cage-reared birds should be of similar age.

### **Preparation of doses**

16. The test substance should be dissolved or suspended in a suitable vehicle or administered in a capsule. If the substance is dissolved or suspended it is recommended that, whenever possible, the use of an aqueous solution or suspension be considered first, followed by consideration of a solution or emulsion in oil (e.g. corn oil) and then by possible solution in other vehicles. For vehicles other than water, the toxicity of the vehicle must be known and should not have emetic properties. The dose will be determined based on bodyweight measured within 24 hours of dosing.

## **PROCEDURE**

### **Limit test**

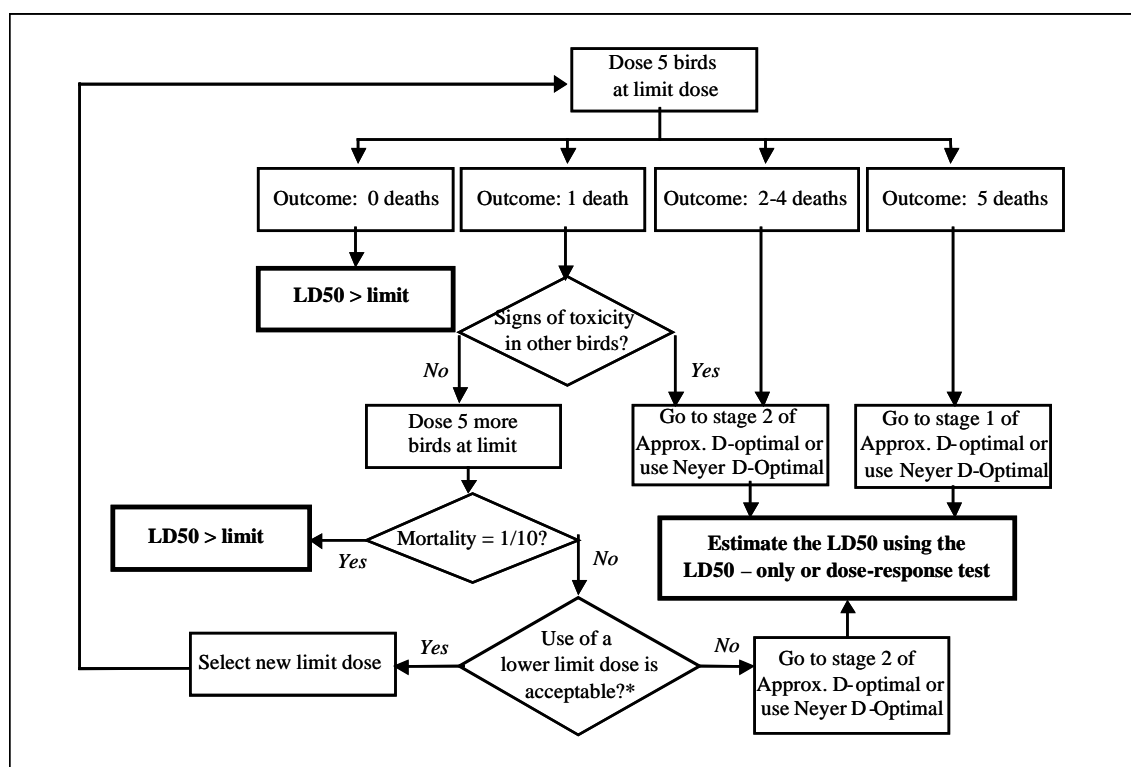
17. The limit test should be employed unless available information suggests that the LD50 is below the limit dose (typically considered to be 2000 mg/kg unless a higher dose is necessary to meet risk criteria). The limit test design is simple, and consists of dosing 5 animals at the limit dose. Birds are given a single oral dose of the test substance (mg/kg bodyweight) into the crop or proventriculus, and then observed for 14 days.

18. Figure 1 describes the procedure to be followed according to the mortality observed. If no mortality occurs, it can be concluded at the 95% confidence level, that the LD50 is above the limit dose. If mortality is observed, several possible options can be considered (see Fig. 1), depending on the number of animals dying and on observed signs of toxicity. More animals can be dosed at the limit, to demonstrate with adequate confidence (i.e. 95%) that the LD50 is above the limit dose, or additional dose levels can be added to the experiment to allow estimation of the LD50 (and slope and confidence limits if required). If only one mortality is observed, and no signs of toxicity are observed in other birds, then five more birds can be dosed at the limit, to attempt to demonstrate with adequate confidence (i.e. 95%), that the LD50 is above the limit dose. If there is only one death in the total of 10 birds, then it can be concluded that at the 95% confidence level, the LD50 is above the limit dose. If there is more than one death in the total of 10 birds, then the limit test can be repeated at a lower limit if this is compatible with risk assessment needs or the approximate D-optimal (starting with stage 2) or Neyer D-optimal design described below can be performed.

19. If the observed mortality is 1 out of 5 birds and there are signs of toxicity in other birds, or if there are between 2 and 4 mortalities among 5 birds, use either the approximate D-optimal (starting with stage 2) or the Neyer D-optimal designs described below. If mortality is complete (i.e. all 5 birds have died), use either the approximate D-optimal (starting with stage 1) or Neyer D-optimal designs described below. Additional dosing can begin before the 14 day period of observation is complete.

## PROPOSAL B

**Figure 1.** Limit Test procedure



20. To proceed from the limit test to stage 2 of the approximate D-optimal method, an initial estimate of the LD50 is needed to select doses at this stage. Appropriate estimates for this use are given in Table 1. Using the appropriate estimate of the LD50, the lowest and highest doses can be calculated as described in the directions for stage 2. The calculated high dose (*hdose*) will generally need to be adjusted (reduced) to the highest level that is consistent with practicality and animal welfare constraints. It should be noted that in some circumstances, the LD50 cannot be estimated without using doses above the limit dose. To use the Neyer D-optimal method an estimate of the slope is needed. In addition, if the outcome of the limit test is complete mortality, an estimate of the lower bound of the LD50 is also required. Because there are constraints on the use of very high doses of test substance, it may not always be possible to estimate the LD50 for slightly toxic substances.

**Table 1.** Initial estimate of LD50 for use in stage 2 of the approximate D-optimal design derived from mortality in a limit test.

Mortality (%)	10	20	30	40	50	60	70	80	90
Est. LD50	3609	2947	2546	2247	2000	1780	1571	1357	1108

### LD50-only test and dose-response test

21. The LD50-only and the dose-response tests both employ staged test designs where animals are dosed at different times (stages) to allow the use of accumulated information to optimally position doses along the dose response curve. The LD50-only test and the dose-response test differ primarily in the total number of birds used, and in the number of dosing stages.

## PROPOSAL B

22. At each stage in both tests, one or more birds are given a single oral dose (mg/kg bodyweight) of the test substance into the crop or proventriculus. All birds are observed for 14 days, but selection of doses for subsequent stages is typically based on results after 3 days. This interval may be reduced if birds quickly show signs of recovery or extended if delayed mortality is expected or observed. If delayed mortality occurs in early stages after dose selection for second or third stages, placement of doses at later stages will not be optimal. In the 1<sup>st</sup> dosing stage, each bird is given a different dose, selected to bracket the best available estimate of the LD50 (e.g. based on the rat or other bird species' LD50). Depending on the outcome of the 1<sup>st</sup> stage, the doses for the 2<sup>nd</sup> stage are determined (see following section). The process continues to a 3<sup>rd</sup> stage in the dose-response test or if added precision is needed in the estimate of the LD50. Observations of deaths that are clearly not treatment related, (e.g. physical injury) should be excluded from calculations.

23. The adequacy of these tests depends on the selection of dose levels, number of doses, number of birds treated at each dose level, and the number of test stages. Acceptable experimental designs include approximate D-optimal designs and Neyer D-optimal designs described below. These designs have been shown to have very desirable performance properties for estimating LD50s, confidence intervals, and dose response slopes. However, practical constraints limit the number of stages that are desirable, and the following descriptions take this into account. Thus, the recommended specifications for the LD50-only test call for at least two dosing stages. The dose-response test is essentially an extension of this test to three or more stages. These designs can be modified if it can be shown that the modified design performs as well as or better than the specified designs.

### The Approximate D-Optimal Design

24. The approximate D-optimal design consists of a number of stages (periods during an experiment in which a number of birds are dosed simultaneously and observed for a period of time), and it is completely flexible with respect to number of stages, and number of birds per stage. If an LD50 estimate only is required, a two-stage twelve-bird design, with four birds in the first stage and eight birds in the second stage, is recommended. Alternatively, if both LD50 and slope estimates are required and/or a confidence interval for the LD50 is required, then more birds should be used. In this case, a three-stage, twenty-four-bird design, with four birds in the first stage and ten birds in the other two stages, is recommended. A detailed explanation of the logic and statistical basis of approximate D-optimal designs is given in Annex 2.

25. Three stages of the approximate D-optimal design are described below. See Annex 3 for a worked example of the calculations. The following parameters are used in this paragraph.

- ldose* -- the lowest dose used during a particular stage (mg active ingredient/kg body weight).
- hdose* -- the highest dose used during a particular stage (mg active ingredient/kg body weight).
- step* -- the multiplication factor used in calculating the individual doses.
- $k_1$  -- the number of doses in stage 1. (typically 4)
- $k_2$  -- the number of doses in stage 2.

**Stage 1:** A series of  $k_1$  doses is equally spaced on a log scale around the initial estimate of the LD50.

(1) Calculate:  $ldose = 0.1414 \times LD50$  and  $hdose = 7.071 \times LD50$ .

If *hdose* is greater than 3330 then set  $hdose = 3330$  (may be less if limited by physical constraints) and recalculate the lowest dose as  $ldose = hdose / 50$ .

(2) Calculate:  $step = 50^{1/(k_1-1)}$

(3) Calculate  $k_1$  doses:  $dose\ i = ldose \times step^{(i-1)}$ , for  $i = 1$  to  $k_1$ ,

(4) Each of the  $k_1$  doses of the test substance calculated in 3 is given to a single bird.

## **PROPOSAL B**

- (5) Dosing is followed by an observation period, at the end of which it is noted whether each bird is dead or alive. An approximate LD50 is computed as the geometric mean of the doses that produced a transition from survival to death. (Table 2)

**Stage 2:** A series of  $k_2$  doses is equally spaced on a log scale around the interim estimate of the LD50 obtained from stage 1.

- (1) Calculate:  $ldose = 0.3425 \times LD50$  and  $hdose = 2.919 \times LD50$ . If  $hdose$  is greater than 3330 then set  $hdose = 3330$  (may be less if limited by physical constraints).
- (2) Calculate:  $step = (hdose / ldose)^{1/(k_2-1)}$
- (3) Calculate the  $k_2$  doses:  $dose_i = ldose \times step^{(i-1)}$ , for  $i = 1$  to  $k_2$ .
- (4) Each dose of the test substance calculated in 3 is given to a single bird.
- (5) Dosing is followed by an observation period, at the end of which it is noted whether each bird is dead or alive. A probit model is fitted to the combined data from stages one and two, and both the LD50 and slope are estimated.

**Stage 3:** Two doses are equally spaced on a log scale around the estimate of the LD50 obtained from stage 2.

- (1) Calculate:  $ldose = 10^{(-1.036 / slope)} \times LD50$  and  $hdose = 10^{(1.036 / slope)} \times LD50$ .  
If an estimate of the slope is not available from stage two it should be assumed to be 5. If the estimate of slope from stage two is either greater than fifteen or less than one, then it should be set to fifteen or one respectively. If  $hdose$  is greater than 3330 then set  $hdose = 3330$  (may be less if limited by physical constraints). Note that in this stage, multiple birds will typically be dosed at each of the two dose levels.
- (2) Dosing is followed by an observation period, at the end of which it is noted whether each bird is dead or alive. A probit model is fitted to the combined data from stages one, two and three, and both the LD50 and slope are estimated.
- (3) The study can be extended by adding additional stages, performed in the same way as stage three, to obtain higher levels of precision in estimates.

## PROPOSAL B

**Table 2.** Example of how the approximate LD50 is computed from the results of stage 1 with 4 doses by computing the geometric mean of the doses that produced a transition from survival to death. Survival is represented by O and death by X.

dose1	dose2	dose3	dose4	Approx. LD50 Estimate
O	O	O	O	$(\text{dose4} \times \text{dose5})^{1/2}$
O	O	O	X	$(\text{dose3} \times \text{dose4})^{1/2}$
O	O	X	O	$(\text{dose2} \times \text{dose3} \times \text{dose4} \times \text{dose5})^{1/4} = (\text{dose3} \times \text{dose4})^{1/2}$
O	X	O	O	$(\text{dose1} \times \text{dose2} \times \text{dose4} \times \text{dose5})^{1/4} = \text{dose3}$
X	O	O	O	$(\text{dose0} \times \text{dose1} \times \text{dose4} \times \text{dose5})^{1/4} = (\text{dose2} \times \text{dose3})^{1/2}$
O	O	X	X	$(\text{dose2} \times \text{dose3})^{1/2}$
O	X	X	O	$(\text{dose1} \times \text{dose2} \times \text{dose4} \times \text{dose5})^{1/4} = \text{dose3}$
X	X	O	O	$(\text{dose0} \times \text{dose1} \times \text{dose4} \times \text{dose5})^{1/4} = (\text{dose2} \times \text{dose3})^{1/2}$
O	X	O	X	$(\text{dose1} \times \text{dose2} \times \text{dose3} \times \text{dose4})^{1/4} = (\text{dose2} \times \text{dose3})^{1/2}$
X	O	X	O	$(\text{dose0} \times \text{dose1} \times \text{dose2} \times \text{dose3} \times \text{dose4} \times \text{dose5})^{1/6} = (\text{dose2} \times \text{dose3})^{1/2}$
X	O	O	X	$(\text{dose0} \times \text{dose1} \times \text{dose3} \times \text{dose4})^{1/4} = \text{dose2}$
O	X	X	X	$(\text{dose1} \times \text{dose2})^{1/2}$
X	O	X	X	$(\text{dose0} \times \text{dose1} \times \text{dose2} \times \text{dose3})^{1/4} = (\text{dose1} \times \text{dose2})^{1/2}$
X	X	O	X	$(\text{dose0} \times \text{dose1} \times \text{dose3} \times \text{dose4})^{1/4} = \text{dose2}$
X	X	X	O	$(\text{dose0} \times \text{dose1} \times \text{dose4} \times \text{dose5})^{1/4} = (\text{dose2} \times \text{dose3})^{1/2}$
X	X	X	X	$(\text{dose0} \times \text{dose1})^{1/2}$

**Note.** Even though only 4 doses (dose1 through to dose 4) were used in the test, values for dose0 and dose5 are mentioned in the table. The values that should be used for these doses are one step up or down from the actual test doses. That is,  $\text{dose0} = \text{dose1} / \text{step}$  and  $\text{dose5} = \text{dose4} \times \text{step}$ . Dose0 must be added to the computation of the approximate LD50 when mortality occurs at the lowest test dose, and dose5 is added when there is survival at the highest test dose.

### **Neyer D-Optimal Design**

26. Small improvements over the approximate D-optimal designs can be achieved by employing Neyer D-Optimal designs (Neyer, 1994). Thus, use of Neyer D-optimal designs is also desirable, although these designs may be more complex than the approximate D-optimal designs. Neyer D-optimal designs require the use of specialized computer software in dose selection (e.g. Sentest©, from Neyer software). These designs are completely flexible with respect to number of doses per stage and number of birds per dose per stage. Best performance is achieved with only one bird per dose in every stage. However, for practical reasons, the number of stages employed should usually be held to 4 or less. The minimum requirements for Neyer D-optimal designs are similar to those for the approximate D-optimal designs. If an LD50 estimate only is required, a twelve bird design with at least two stages is recommended. Alternatively, if in addition to an LD50 estimate, a slope estimate and/or a confidence interval for the LD50 are required, then more birds should be used. In this case, a 3 stage, twenty-four bird design, is recommended. A more detailed description of Neyer D-optimal designs is given in Appendix 1.

### **Administration of dose**

27. The test substance is administered in a single dose by gavage or capsules. The dosing volume must remain constant with respect to bodyweight and should not exceed 10ml/kg bodyweight. Birds should be fasted for 12-15 hours overnight immediately prior to dosing. Shorter fasting periods may be necessary for smaller and wild caught species. Regurgitation must be recorded. The addition of a

## **PROPOSAL B**

non-toxic coloured food dye which contrasts with faeces will allow regurgitation to be more easily recognised.

28. Regurgitation compromises the evaluation of toxicity and is a feature of acute oral toxicity testing in birds. It may be related to the dosing technique or characteristics of the test substance. The frequency of regurgitation may be reduced by lowering the dose volume or by changing carriers.

### **Observations**

29. Birds are observed individually during the first 2 hours after dosing, on at least 3 evenly spread additional occasions during the first 24 hours and at least daily thereafter for a total of 14 days. However, the duration of the observation period should not be fixed rigidly. It should be determined by the cessation of clinical symptoms and death and may thus be extended when considered necessary.

30. Observations on each individual should include regurgitation, signs of intoxication and remission, abnormal behaviour, mortality and time to death.

31. In the absence of an untreated control there is limited value in measuring bodyweight and food consumption. However, birds may be weighed at the start and end of the study to determine weight loss. Food consumption may be measured on days 1, 3, 7 and 14 days after dosing. Gross pathology should be undertaken on all birds from each treatment group to help identify incidental mortalities and obvious symptoms of toxicity.

### **DATA AND REPORTING**

#### **Data**

32. Individual bird data should be provided and summarised in tabular form, showing the dose, number of birds tested, signs of toxicity, death and numbers sacrificed for humane reasons, time of death of individual birds, a description and the time course of toxic effects and reversibility, and gross pathological findings.

#### **Calculation of the LD50**

33. The methods for estimating the LD50 (median lethal dose) are the same for staged tests as for other types of dose response tests. Certain features of the staged designs are based on the assumptions that the underlying form of the dose response curve approximates the probit model, so clearly, the use of a probit regression model (with the logarithm of dose as the independent variable) to estimate the LD50 is appropriate. For the probit model, the maximum likelihood estimate of the  $\text{Log}(\text{LD50})$  is  $-a/b$ , where  $a$  is the intercept and  $b$  is the slope. Probit regression is fully integrated into available software used to support Neyer D-optimal study designs. The logistic model is very similar in shape to the probit model, and can be used in its place in the approximate D-optimal design. Furthermore, when the mortality data from all stages of a study is pooled for analysis, other models may also be appropriate if the probit model fails to fit the observations.

34. Standard maximum likelihood methods used to fit models to the data can be used only if there are two treatments that display partial mortality (i.e. neither 0 nor 100%), or there is a reversal in trend (i.e. mortality at a lower dose is greater than mortality in one of the higher doses) somewhere in the observed responses. Occasionally, these conditions may not be obtained. In such cases, various interpolation and moving average methods may be used to estimate the LD50.

35. Confidence intervals for the LD50 can be obtained using Fieller's theorem, (Fieller 1940) likelihood ratio methods (Crump and Howe, 1985, Piegorsch and Bailer 1997), or by binomial



## **PROPOSAL B**

methods (Chapman et. al. 2002). Many standard statistical packages have built in routines to estimate the LD50 and its confidence interval (SAS©, LogXact©, Toxstat©, BMDS [available from USEPA - calculates lower confidence limit only]).

36. It should be recognized that it is often possible to calculate the slope and confidence intervals for the LD50 from the results of the LD50-only test. However, due to the small number of replicates, these estimates may be unreliable. Thus, the slope and confidence intervals for the LD50 should be reported only for the full dose-response test. Some measure of goodness of fit of the regression model (e.g. Pearson goodness of fit, likelihood ratio Chi-square) should be reported for each dose-response test. Reporting of a measure of goodness of fit is optional for the LD50-only test.

### **Report**

37. The report should contain the following minimum information to confirm compliance with the Guideline and test results:

#### test substance

- identification
- batch and lot number
- purity
- stability at room temperature
- volatility

#### test method and system

- test type
- test species, source, strain, age, weight, health
- description of test method

#### conduct of test

- test groups and design (no treatments and replicates, individual or group caging)
- acclimation and assignment procedures (duration, randomisation)
- dose method (gavage/capsule, carrier/solvent, volume/bird as % bodyweight)
- housing conditions (type, size, pen materials, floor covering, temperature, humidity, photoperiod, light intensity)
- food and water (availability, identification, source, composition, calorific value, results of contaminant analysis)
- frequency, duration and method of observations (health/mortality, bodyweight, food consumption)
- description of statistical methods

#### results of test

- mortality (time to death, clinical symptoms, calculation of LD<sub>50</sub>)
- gross pathological examination

### **LITERATURE**

- (1) OECD 1996. Report of the SETAC/OECD Workshop on Avian Toxicity Testing (1996). OECD/GD (96) 166.
- (2) Chapman, P.F., T. A. Springer, R. Dark, B. T. Collins, D. Farrar, J. W. Green, G. F. Healy, B.T. Neyer, W. Slob, . 2002. Avian Acute Oral Toxicity: A Statistical Evaluation of Sequential Experimental Designs. In preparation.

## **PROPOSAL B**

- (3) Crump, K.S. and R. Howe. 1985. A review of methods for calculating confidence intervals in low dose extrapolation. In: D. Krewski, Ed. Toxicology and Risk Assessment. CRC Press. Canada.
- (4) Efron, B. and R. Tibshirani. 1993 Introduction to the bootstrap. Chapman and Hall. New York.
- (5) Neyer, B.T. 1994. A D-Optimality Based Sensitivity Test, Technometrics, 36:61-70.
- (6) Piegorsch, W. W. and A.J. Bailer. 1997. Statistics for environmental biology and toxicology. Chapman and Hall. London.

# PROPOSAL B

## ANNEX 1

### DEFINITIONS

Acute oral toxicity refers to those adverse effects occurring following oral administration of a single dose of a substance, or multiple doses given within 24 hours.

Dose is the amount of test substance administered. Dose is expressed as weight (g, mg) or as weight of test substance per unit weight of test animal (e.g. mg/kg).

D-Optimality refers to a design that optimises the estimates of the LD50 and the slope at the same time. If we wish to jointly estimate two parameters from a statistical model, as is the case with a probit model, then we can present the confidence limits in a two dimensional graph as an ellipse surrounding the estimates. The D-optimality criterion assigns doses in such a way that the area of the confidence ellipse is minimised. Thus D-optimality treats estimates of both LD50 and slope as being of equal importance. The number of D-optimal dose levels is equal to the number of parameters in the statistical model fitted to the data. For the probit model this means two dose levels. (In order to use D-optimality we need to know the true values for the parameters in the statistical model, which of course we never do. However, in a sequential design we can use parameter estimates from earlier stages to derive approximate D-optimal doses for the current stage.)

LD50 (median lethal oral dose), is a statistically derived single dose of a substance that can be expected to cause death in 50 per cent of animals when administered by the oral route. The LD50 value is expressed in terms of weight of test substance per unit weight of test animal (mg/kg).

Limit dose refers to a dose at an upper limitation on testing.

Phase (in the approximate D-optimal and Neyer D-optimal designs) refers to a number of stages, each of which may have similar designs, although stages from different phases will usually have different designs. In order to move from one phase to another, a “stopping criterion” of some sort needs to be satisfied. The criterion may simply be completing a fixed number of stages, or it may be a satisfying a condition of some sort, that can be met after a variable number of test stages.

Probit is an abbreviation for the term “probability integral transformation” and a probit dose-response model permits a standard normal distribution of expected responses (i.e., one centered to its mean and scaled to its standard deviation, *sigma*) to doses (typically in a logarithmic scale) to be analyzed as if it were a straight line with slope the reciprocal of *sigma*. A standard normal lethality distribution is symmetric; hence, its mean is also its true LD50 or median response.

Slope (of the dose-response curve) is a value related to the angle at which the dose response curve rises from the dose axis. In the case of probit analysis, when responses are analyzed on a probit scale against dose on a log scale this curve will be a straight line and the slope is the reciprocal of *sigma*, the standard deviation of the underlying test subject tolerances, which are assumed to be normally distributed. See probit and *sigma*.

Stage (in the approximate D-optimal and Neyer D-optimal designs) refers to a period during an experiment in which a number of birds are dosed simultaneously, and observed for a period of time. Provisional estimates of LD50 and slope from the previous stage may influence the design of the stage. Likewise the data resulting from the stage may influence the design of succeeding stages.

## **PROPOSAL B**

### **ANNEX 2**

#### **DETAILED DESCRIPTIONS OF THE APPROXIMATE D-OPTIMAL AND NEYER D-OPTIMAL DESIGNS**

##### **TOLERANCE DISTRIBUTIONS AND THE PROBIT MODEL**

1. The philosophy underlying tolerance distributions is that an individual bird will die if it receives a dose above a certain value and will survive if the dose is less than this value. The specific value is called a tolerance and is assumed to be fixed for an individual bird, but to vary between birds. Thus if we have a population of birds we can speak of a distribution of tolerances, or a tolerance distribution.

2. In order to estimate the tolerance distribution from a sample of birds we fit a statistical model. If we assume that the tolerances follow a normal distribution we fit a probit model which takes the form:

$$\text{Probit}(p) = \alpha + \beta \cdot \log(d)$$

Where:  $p$  is the probability that the tolerance of an individual bird is less than dose  $d$  – i.e. the probability that a bird receiving dose  $d$  will survive.  
 $\alpha$  and  $\beta$  are parameters representing the intercept and slope of a straight line relationship between probit ( $p$ ) and  $\log(d)$

3. The probit model is fitted to test data in order to obtain estimates of the parameters  $\alpha$  and  $\beta$  which we call  $a$  and  $b$  respectively. The estimate of the mean of the tolerance distribution of the population of birds (called the LD50) can then be determined from the equation:

$$\text{Estimate } (\log(\text{LD50})) = -a/b,$$

And the variance of the tolerance distribution can be estimated by:

$$\text{Variance} = 1/b.$$

4. A variety of techniques are available for estimating the confidence limits for estimates of  $\alpha$ ,  $\beta$ ,  $\log(\text{LD50})$  and the variance of the tolerance distribution.

5. Once the tolerance distribution has been estimated (i.e. once estimates of  $\alpha$  and  $\beta$  have been obtained) percentiles of the distribution can be estimated (dose levels corresponding to specific values of  $p$ ).

##### **APPROXIMATE D-OPTIMAL MULTI-STAGE DESIGNS**

6. This design consists of one phase but is partitioned into a number of stages. It is completely flexible with respect to number of stages, number of doses per stage, and number of birds per dose. At the start of the test it is assumed that we can make some initial estimate of the value the LD50. In practice this estimate may be made using knowledge of the chemistry and toxicity of similar active ingredients, and also from the results of tests on other bird species or mammals. The test generally has at least two stages, and may be extended to three or more. The first three stages differ considerably from each other. The first is a 'ranging' stage designed to confirm and improve an initial guess at the LD50. The second is designed to accurately locate the LD50. The third (and any subsequent stages)

## **PROPOSAL B**

are intended to provide estimates of the confidence interval and slope of the dose response curve or to further reduce uncertainty in the estimate of the LD50.

7. In the first stage the doses are equally spaced on a log scale around the initial estimate of the LD50 and the ratio of the highest to the lowest dose is set to 50. (Value of 50 chosen based on extensive simulation by Chapman et. al. 2002) Each dose of the test substance is given to a single bird. An approximate estimate of the LD50 is computed as the geometric mean of the transition doses (see Table 2 in main body of text.). This method of estimating the LD50 has several advantages. First, it can provide a reasonable estimate given very little information, and second, it can be adapted to provide a working estimate of the LD50 when there is mortality at the lowest dose or survival at the highest dose. Either of these cases suggests that the actual LD50 may be outside of the range of the test concentrations. It is then assumed that if there was not survival at the low dose, there would be at a transition to survival at low dose/step. The geometric mean of transitions, including these values is then taken as the estimate the LD50. Likewise if there was no mortality at the high dose, it is assumed that there would be at a transition to mortality at high dose\*step . The geometric mean of transitions, including these values is then taken as the estimate the LD50.

8. In stage 2 the doses are equally spaced on the log scale about the LD50 estimate computed in stage 1. The extreme doses are placed at those points corresponding to 1 and 99 percent kill, assuming a slope of 5. A probit model is fitted to the combined data from stages 1 and 2 and both the LD50 and slope are estimated. A review of the EPA One-Liner database suggests that for the historical record of pesticide tests, the modal slope for avian acute tests is 5.

9. In stage 3 the choice of doses is based on estimates of both the LD50 and slope from stage 2. If an estimate of the slope is not available from stage 2 it is assumed to be 5. If the estimate of slope from stage 2 is either greater than fifteen or less than one, then it is set to be fifteen or one respectively. These limits on slope estimates are needed because maximum likelihood estimates using small numbers of observations may give 'wild' (probably grossly inaccurate) slope estimates. Only two doses are used, each at one of the D-optimal doses: half of the birds are given a lower dose corresponding to 15% kill, and the remaining birds are given an upper dose corresponding to 85% kill. A probit model is fitted to data from the current and previous stages and an estimate of the LD50 and slope is obtained.

10. All additional stages are identical to stage 3.

### **NEYER D-OPTIMAL MULTI-STAGE DESIGNS**

11. This design consists of four phases, with each phase (potentially) consisting of a number of stages. It is completely flexible with respect to number of doses per phase and number of birds per dose per phase. However this description considers designs that use only one bird per dose in every stage of the first three phases and only two doses in stages in the phase 4. At the start of the test it is assumed that realistic upper and lower bounds for the LD50 can be determined and that we have some knowledge of the slope. This initial knowledge may come from a variety of sources.

12. In stage 1 of phase 1 the doses are equally spaced on a log scale between the upper and lower bounds. If at the end of the period of observation all birds have survived or all have died then another stage is run. Otherwise phase 1 finishes. When they are run, all additional stages are identical to the first except the bounds are changed. Following complete survival the upper bound is increased by the difference between the upper and lower bound used in all previous stages. Thus, the range doubles in size at each stage after the second. Similarly, following complete mortality the lower bound is

## **PROPOSAL B**

decreased in the same fashion. Phase 1 ends when at least one dose results in mortality and at least one result in survival.

13. In every stage of phase 2 the doses are equally spaced on a log scale between the highest dose resulting in survival and the lowest dose resulting in mortality. The doses chosen will not include highest dose resulting in survival and the lowest dose resulting in mortality. Phase 2 ends when the difference between these doses is less than the reciprocal of the initial estimate of the slope (i.e. 1 divided by the initial estimate of the slope). If at the end of a stage the lowest dose resulting in mortality is larger than the highest dose resulting in survival the algorithm proceeds straight to phase 4. If at the start of phase 2 the difference between the highest dose resulting in survival and the lowest dose resulting in mortality is less than the reciprocal of the initial estimate of the slope, phase 2 is skipped and the algorithm proceeds straight to phase 3 or 4.

14. At the beginning of phase 3 it is assumed that the true LD50 is equal to the geometric mean of the highest dose resulting in survival and the lowest dose resulting in mortality and that the true slope is equal to the initial estimate of the slope. The results of all previous stages are used in the calculations. These values are then used to compute the two D-optimal dose levels. In stage 1 the doses are equally spaced on a log scale between the two d-optimal dose levels, including the end points. If at the end of any stage the lowest dose resulting in mortality is larger than the highest dose resulting in survival then phase 3 ends. Otherwise another stage is run. All additional stages, when run, are identical, except that the reciprocal of the initial estimate of the slope is decreased by twenty percent every time it is used (i.e. the slope estimate is increased by twenty-five percent). When this reduction causes the difference between the highest dose resulting in survival and the lowest dose resulting in mortality to be larger than the reciprocal of the slope the algorithm returns to phase 2. At the end of phase 3 a probit model is fitted to data from all previous stages in order to derive an estimate of the LD50 and the slope.

15. Phase 4 consists of an arbitrary number of stages. For stage 1 estimates of the LD50 and slope obtained at the end of phase 3 are used to determine D-optimal dose levels. The birds dosed in stage 1 are then divided between these two levels in such a way as to balance the overall number of birds tested at the two doses. The two D-optimal dose levels correspond approximately to the 13% and 87% points of the tolerance distribution curve. At the end of stage 1 a probit model is fitted to data from all previous stages in order to compute an estimate of the LD50 and the slope. When more than one stage is used in phase 4 the D-optimal dose levels are determined afresh for each stage using the most recent estimate of the LD50 and slope derived at the end of the previous stage. At the end of phase 4 a probit model is fitted to data from all previous stages in order to derive an estimate of the LD50 and the slope.

## **PROPOSAL B**

### **ANNEX 3**

#### **APPROXIMATE D-OPTIMAL DESIGN - WORKED EXAMPLE**

1. In this appendix the calculations involved in a three stage, twenty-four bird approximate D-optimal design are illustrated. This design is recommended if both LD50 and slope estimates are required and/or the confidence interval for the LD50 is required.

2. At the start of the test, the initial guess for the LD50 is 100.

3. **Stage 1:** Four doses are equally spaced on a log scale around the initial guess of the LD50. The doses can be calculated by following steps 1-3 in the guideline:

(1)  $ldose = 0.1414 \times 100 = 14.14$  and  $hdose = 7.071 \times 100 = 707$ .

(2)  $step = 50^{1/3} = 3.684$

(3) Thus the doses are given by:  $dose_i = 14.14 \times 3.684^{(i-1)}$ , for  $i = 1$  to 4. Using this equation the four doses are: 14.10, 52.09, 191.91, and 707.

4. Each of the four doses calculated in 3 is given to a single bird. At the end of the observation period we find that the bird given dose 14.10 is alive, and the other three birds are dead. An approximate LD50 is computed as the geometric mean of the doses that produced a transition from survival to death, i.e.

$$LD50 = (14.14 \times 52.09)^{1/2} = 27.14 \text{ (see Table 2 for details).}$$

5. **Stage 2:** Ten doses are equally spaced on a log scale around the estimate of the LD50 obtained from stage 1. The doses can be calculated by following steps 1-3 in the guideline.

(1)  $ldose = 0.3425 \times 27.14 = 9.30$  and  $hdose = 2.919 \times 27.14 = 79.23$

(2)  $step = (79.23/9.30)^{1/9} = 1.269$

(3) Thus the doses are given by:  $dose_i = 9.30 \times 1.269^{(i-1)}$ , for  $i = 1$  to 10. Using this equation, the ten doses are: 9.30, 11.80, 14.98, 19.00, 24.12, 30.60, 38.84, 49.28, 62.54, and 79.23.

6. Each of the ten doses calculated in 3 is given to a single bird. At the end of the observation period the birds given doses 9.30, 11.80, 14.98, 19.00, 24.12, and 30.60 are alive, and the other four birds are dead. We fit a probit model to the combined data from stages 1 and 2 in SAS, and obtain the following estimates: LD50 = 34.48 and slope = 162.57.

7. **Stage 3:** Two doses are equally spaced on a log scale around the estimate of the LD50 obtained from stage 2. The two doses,  $ldose$  and  $hdose$ , can be calculated following step 1 in the guideline. Because the estimate of the slope from stage 2 is greater than fifteen it is set to fifteen. Then:

(1)  $ldose = 10^{(-1.036/15)} \times 34.48 = 29.41$  and  $hdose = 10^{(1.036/15)} \times 34.48 = 40.42$ .

8. Each of the two doses calculated in 1 is given to five birds. At the end of the observation period two birds have died at the lowest dose and four birds have died at the highest dose. We fit a probit model to the combined data from stages 1, 2 and 3 using PROC PROBIT in the SAS statistical software system and obtain the following results: LD50 = 32.67, with 95% confidence interval (21.61, 41.53), and slope = 11.31, with 95% confidence interval (1.50, 21.12).