APPENDIX B

RELEVANT FEDERAL ACUTE ORAL TOXICITY REGULATIONS AND TESTING GUIDELINES

[This Page Intentionally Left Blank]

APPENDIX B1

TABLE OF RELEVANT ACUTE ORAL TOXICITY REGULATIONS

(Note to the Reader: Regulations may be updated in the future. It is recommended that users review the most current version of all regulations identified. Electronic versions of the regulations can be obtained at: http://www.gpoaccess.gov/nara/index.html)

[This Page Intentionally Left Blank]

AGENCY	TITLE	CHAPTER	PART AND TITLE	SECTION	
CPSC	16	II	PART 1500--HAZARDOUS SUBSTANCES AND ARTICLES; ADMINISTRATION AND ENFORCEMENT REGULATIONS	1500.3	Definitions.
DOT	49	I	PART 173--SHIPPERS--GENERAL REQUIREMENTS FOR SHIPMENTS AND PACKAGINGS	173.132	Class 6, Division 6.1 – Definitions.
				173.133	Assignment of Packing Group and Hazard Zones for Divusion 6.1 Materials.
EPA	40	I	PART 156--LABELING REQUIREMENTS FOR PESTICIDES AND DEVICES	156.10	Labeling Requirements.
				156.620	Toxicity Category.
EPA	40	1	157: PACKAGING REQUIREMENTS FOR PESTICIDES AND DEVICES	157.22	When required.
EPA	40	1	158: DATA REQUIREMENTS FOR REGISTRATION	158.202	Purposes of the registration data requirements.
				158.340	Toxicology data requirements.
				158.690	Biochemical pesticides data requirements.
				158.740	Microbial pesticides--Product analysis data requirements.
EPA	40	I	159: STATEMENTS OF POLICIES AND INTERPRETATIONS	159.165	Toxicological and ecological studies.
OSHA	29	XVII	1910: OCCUPATIONAL SAFETY AND HEALTH STANDARDS	1910.1200	Hazard communication.

Abbreviations: CPSC=U.S. Consumer Products Safety Commission; DOT=Department of Transportation; EPA=U.S. Environmental Protection Agency; OSHA=Occupational Safety and Health Administration.

[This Page Intentionally Left Blank]

APPENDIX B2

OECD GUIDELINE 425: ACUTE ORAL TOXICITY – UP-AND-DOWN PROCEDURE

[This Page Intentionally Left Blank]

OECD/OCDE **425**

Adopted:
17th December 2001

OECD GUIDELINE FOR TESTING OF CHEMICALS

Acute Oral Toxicity – Up-and-Down Procedure

INTRODUCTION

1. OECD guidelines for the Testing of Chemicals are periodically reviewed in the light of scientific progress or changing assessment practices. The concept of the up-and-down testing approach was first described by Dixon and Mood (1)(2)(3)(4). In 1985, Bruce proposed to use an up-and-down procedure (UDP) for the determination of acute toxicity of chemicals (5). There exist several variations of the up-and-down experimental design for estimating an LD50. This guideline is based on the procedure of Bruce as adopted by ASTM in 1987 (6) and revised in 1990. A study comparing the results obtained with the UDP, the conventional LD50 test and the Fixed Dose Procedure (FDP, Guideline 420) was published in 1995 (7). Since the early papers of Dixon and Mood, papers have continued to appear in the biometrical and applied literature, examining the best conditions for use of the approach (8)(9)(10)(11). Based on the recommendations of several expert meetings in 1999, an additional revision was considered timely because: i) international agreement had been reached on harmonised LD50 cut-off values for the classification of chemical substances, ii) testing in one sex (usually females) is generally considered sufficient, and iii) in order for a point estimate to be meaningful, there is a need to estimate confidence intervals (CI).

2. The test procedure described in this Guideline is of value in minimizing the number of animals required to estimate the acute oral toxicity of a chemical. In addition to the estimation of LD50 and confidence intervals, the test allows the observation of signs of toxicity. Revision of Test Guideline 425 was undertaken concurrently with revisions to the Test Guidelines 420 and 423.

3. Guidance on the selection of the most appropriate test method for a given purpose can be found in the Guidance Document on Oral Toxicity Testing (12). This Guidance Document also contains additional information on the conduct and interpretation of Guideline 425.

4. Definitions used in the context of this Guideline are set out in Annex 1.

INITIAL CONSIDERATIONS

5. The testing laboratory should consider all available information on the test substance prior to conducting the study. Such information will include the identity and chemical structure of the test substance; its physical chemical properties; the results of any other *in vitro* or *in vivo* toxicity tests on the substance; toxicological data on structurally related substances or similar mixtures; and the anticipated use(s) of the substance. This information is useful to determine the relevance of the test for the protection of human health and the environment, and will help in the selection of an appropriate starting dose.

6. The method permits estimation of an LD50 with a confidence interval and the results allow a substance to be ranked and classified according to the Globally Harmonised System for the classification of chemicals which cause acute toxicity (16).

7. When no information is available to make a preliminary estimate of the LD50 and the slope of the dose-response curve, results of computer simulations have suggested that starting near 175 mg/kg and using half-log units (corresponding to a dose progression of factor 3.2) between doses will produce the best results. This starting dose should be modified if the substance is likely to be highly toxic. The half-log spacing provides for a more efficient use of animals, and increases accuracy in the prediction of the LD50 value. Because the method has a bias toward the starting dose, it is essential that initial dosing occur below the estimated LD50. (See paragraphs 32 and Annex 2 for discussion of dose sequences and starting values). However, for chemicals with large variability (i.e., shallow dose-response slopes), bias can still be introduced in the lethality estimates and the LD50 will have a large statistical error, similar to other acute toxicity methods. To correct for this, the main test includes a stopping rule keyed to properties of the estimate rather than a fixed number of test observations (see paragraph 33).

8. The method is easiest to apply to materials that produce death within one or two days. The method would not be practical to use when considerably delayed death (five days or more) can be expected.

9. Computers are used to facilitate animal-by-animal calculations that establish testing sequences and provide final estimates.

10. Test substances, at doses that are known to cause marked pain and distress due to corrosive or severely irritant actions, need not be administered. Moribund animals or animals obviously in pain or showing signs of severe and enduring distress shall be humanely killed, and are considered in the interpretation of the test results in the same way as animals that died on test. Criteria for making the decision to kill moribund or severely suffering animals, and guidance on the recognition of predictable or impending death are the subject of a separate OECD Guidance Document (13).

11. A limit test can be used efficiently to identify chemicals that are likely to have low toxicity.

PRINCIPLE OF THE LIMIT TEST

12. The Limit Test is a sequential test that uses a maximum of 5 animals. A test dose of 2000, or exceptionally 5000 mg/kg, may be used. The procedures for testing at 2000 and 5000 mg/kg are slightly different (see paragraphs 23-25 for limit test at 2000 mg/kg and paragraphs 26-30 for limit test at 5000 mg/kg). The selection of a sequential test plan increases the statistical power and also has been made to intentionally bias the procedure towards rejection of the limit test for compounds with LD50s near the limit dose; i.e., to err on the side of safety. As with any limit test protocol, the probability of correctly classifying a compound will decrease as the actual LD50 more nearly resembles the limit dose.

PRINCIPLE OF THE MAIN TEST

13. The main test consists of a single ordered dose progression in which animals are dosed, one at a time, at a minimum of 48-hour intervals. The first animal receives a dose a step below the level of the best estimate of the LD50. If the animal survives, the dose for the next animal is increased by [a factor of] 3.2 times the original dose; if it dies, the dose for the next animal is decreased by a similar dose progression. (Note: 3.2 is the default factor corresponding to a dose progression of one half log unit). Paragraph 32 provides further guidance for choice of dose spacing factor.) Each animal should be observed carefully for up to 48 hours before making a decision on whether and how much to dose the next animal. That decision is based on the 48-hour survival pattern of all the animals up to that time. (See paragraphs 31 and 35 on choice of dosing interval). A combination of stopping criteria is used to keep the

number of animals low while adjusting the dosing pattern to reduce the effect of a poor starting value or low slope (see paragraph 34). Dosing is stopped when one of these criteria is satisfied (see paragraphs 33 and 41), at which time an estimate of the LD50 and a confidence interval are calculated for the test based on the status of all the animals at termination. For most applications, testing will be completed with only 4 animals after initial reversal in animal outcome. The LD50 is calculated using the method of maximum likelihood (14)(15). (See paragraphs 41 and 43.)

14. The results of the main test procedure serve as the starting point for a computational procedure to provide a confidence interval estimate where feasible. A description of the basis for this CI is outlined in paragraph 45.

DESCRIPTION OF THE METHOD

Selection of animals species

15. The preferred rodent species is the rat although other rodent species may be used. Normally female rats are used (12). This is because literature surveys of conventional LD50 tests show that usually there is little difference in sensitivity between sexes, but in those cases where differences are observed, females are generally slightly more sensitive (7). However, if knowledge of the toxicological or toxicokinetic properties of structurally related chemicals indicates that males are likely to be more sensitive then this sex should be used. When the test is conducted in males, adequate justification should be provided.

16. Healthy young adult animals of commonly used laboratory strains should be employed. Females should be nulliparous and non-pregnant. At the commencement of its dosing, each animal should be between 8 and 12 weeks old and its weight should fall in an interval within ± 20 % of the mean initial weight of any previously dosed animals.

Housing and feeding conditions

17. The temperature in the experimental animal room should be 22°C (± 3°C). Although the relative humidity should be at least 30 % and preferably not exceed 70 % other than during room cleaning the aim should be 50-60%. Lighting should be artificial, the sequence being 12 hours light and 12 hours dark. The animals are housed individually. For feeding, conventional rodent laboratory diets may be used with an unlimited supply of drinking water.

Preparation of animals

18. The animals are randomly selected, marked to permit individual identification, and kept in their cages for at least 5 days prior to dosing to allow for acclimatisation to the laboratory conditions. As with other sequential test designs, care must be taken to ensure that animals are available in the appropriate size and age range for the entire study.

Preparation of doses

19. In general test substances should be administered in a constant volume over the range of doses to be tested by varying the concentration of the dosing preparation. Where a liquid end product or mixture is to be tested, however, the use of the undiluted test substance, i.e., at a constant concentration, may be more relevant to the subsequent risk assessment of that substance, and is a requirement of some regulatory

authorities. In either case, the maximum dose volume for administration must not be exceeded. The maximum volume of liquid that can be administered at one time depends on the size of the test animal. In rodents, the volume should not normally exceed 1 mL/100g of body weight; however in the case of aqueous solutions, 2 mL/100g body weight can be considered. With respect to the formulation of the dosing preparations, the use of an aqueous solution/suspension/emulsion is recommended wherever possible, followed in order of preference by a solution/suspension/emulsion in oil (e.g. corn oil) and then possibly solution in other vehicles. For vehicles other than water the toxicological characteristics of the vehicle should be known. Doses must be prepared shortly prior to administration unless the stability of the preparation over the period during which it will be used is known and shown to be acceptable.

PROCEDURE

Administration of doses

20. The test substance is administered in a single dose by gavage using a stomach tube or a suitable intubation cannula. In the unusual circumstance that a single dose is not possible, the dose may be given in smaller fractions over a period not exceeding 24 hours.

21. Animals should be fasted prior to dosing (e.g., with the rat, food but not water should be withheld overnight; with the mouse, food but not water should be withheld for 3-4 hours). Following the period of fasting, the animals should be weighed and the test substance administered. The fasted body weight of each animal is determined and the dose is calculated according to the body weight. After the substance has been administered, food may be withheld for a further 3-4 hours in rats or 1-2 hours in mice. Where a dose is administered in fractions over a period of time, it may be necessary to provide the animals with food and water depending on the length of the period.

Limit test and main test

22. The limit test is primarily used in situations where the experimenter has information indicating that the test material is likely to be nontoxic, i.e., having toxicity below regulatory limit doses. Information about the toxicity of the test material can be gained from knowledge about similar tested compounds or similar tested mixtures or products, taking into consideration the identity and percentage of components known to be of toxicological significance. In those situations where there is little or no information about its toxicity, or in which the the test material is expected to be toxic, the main test should be performed.

Limit test

Limit test at 2000 mg/kg

23. Dose one animal at the test dose. If the animal dies, conduct the main test to determine the LD50. If the animal survives, dose four additional animals sequentially so that a total of five animals are tested. However, if three animals die, the limit test is terminated and the main test is performed. The LD50 is greater than 2000 mg/kg if three or more animals survive. If an animal unexpectedly dies late in the study, and there are other survivors, it is appropriate to stop dosing and observe all animals to see if other animals will also die during a similar observation period (see paragraph 31 for initial observation period). Late deaths should be counted the same as other deaths. The results are evaluated as follows (O=survival, X=death).

24. The LD50 is less than the test dose (2000 mg/kg) when three or more animals die.

O XO XX
O OX XX
O XX OX
O XX X

If a third animal dies, conduct the main test.

25. Test five animals. The LD50 is greater than the test dose (2000 mg/kg) when three or more animals survive.

O OO OO
O OO XO
O OO OX
O OO XX
O XO XO
O XO OO/X
O OX XO
O OX OO/X
O XX OO

Limit Test at 5000 mg/kg

26. Exceptionally, and only when justified by specific regulatory needs, the use of a dose at 5000 mg/kg may be considered (see Annex 4). For reasons of animal welfare concern, testing of animals in GHS Category 5 ranges (2000-5000mg/kg) is discouraged and should only be considered when there is a strong likelihood that results of such a test have a direct relevance for protecting human or animal health or the environment.

27. Dose one animal at the test dose. If the animal dies, conduct the main test to determine the LD50. If the animal survives, dose two additional animals. If both animals survive, the LD50 is greater than the limit dose and the test is terminated (i.e. carried to full 14-day observation without dosing of further animals).

28. If one or both animals die, then dose an additional two animals, one at a time. If an animal unexpectedly dies late in the study, and there are other survivors, it is appropriate to stop dosing and observe all animals to see if other animals will also die during a similar observation period (see paragraph 10 for initial observation period). Late deaths should be counted the same as other deaths. The results are evaluated as follows (O=survival, X=death, and U=Unnecessary).

29. The LD50 is less than the test dose (5000 mg/kg) when three or more animals die.

O XO XX
O OX XX
O XX OX
O XX X

30. The LD50 is greater than the test dose (5000 mg/kg) when three or more animals survive.

O OO
O XO XO

O XO O
O OX XO
O OX O
O XX OO

Main test

31. Single animals are dosed in sequence usually at 48 h intervals. However, the time intervals between dosing is determined by the onset, duration, and severity of toxic signs. Treatment of an animal at the next dose should be delayed until one is confident of survival of the previously dosed animal. The time interval may be adjusted as appropriate, e.g., in case of inconclusive response. The test is simpler to implement when a single time interval is used for making sequential dosing decisions. Nevertheless, it is not necessary to recalculate dosing or likelihood-ratios if the time interval changes midtest. For selecting the starting dose, all available information, including information on structurally related substances and results of any other toxicity tests on the test material, should be used to approximate the LD50 as well as the slope of the dose-response curve.

32. The first animal is dosed a step below the best preliminary estimate of the LD50. If the animal survives, the second animal receives a higher dose. If the first animal dies or appears moribund, the second animal receives a lower dose. The dose progression factor should be chosen to be the antilog of 1/(the estimated slope of the dose-response curve) (a progression of 3.2 corresponds to a slope of 2) and should remain constant throughout testing. When there is no information on the slope of the substance to be tested, a dose progression factor of 3.2 is used. Using the default progression factor, doses would be selected from the sequence 1.75, 5.5, 17.5, 55, 175, 550, 2000 (or 1.75, 5.5, 17.5, 55, 175, 550, 1750, 5000 for specific regulatory needs). If no estimate of the substance's lethality is available, dosing should be initiated at 175 mg/kg. In most cases, this dose is sublethal and therefore serves to reduce the level of pain and suffering. If animal tolerances to the chemical are expected to be highly variable (i.e., slopes are expected to be less than 2.0), consideration should be given to increasing the dose progression factor beyond the default 0.5 on a log dose scale (i.e., 3.2 progression factor) prior to starting the test. Similarly, for test substances known to have very steep slopes, dose progression factors smaller than the default should be chosen. (Annex 2 includes a table of dose progressions for whole number slopes ranging from 1 to 8 with starting dose 175 mg/kg).

33. Dosing continues depending on the fixed-time interval (e.g., 48-hour) outcomes of all the animals up to that time. The testing stops when one of the following stopping criteria first is met:

(a) 3 consecutive animals survive at the upper bound;

(b) 5 reversals occur in any 6 consecutive animals tested;

(c) at least 4 animals have followed the first reversal and the specified likelihood-ratios exceed the critical value. (See paragraph 44 and Annex 3. Calculations are made at each dosing, following the fourth animal after the first reversal).

For a wide variety of combinations of LD50 and slopes, stopping rule (c) will be satisfied with 4 to 6 animals after the test reversal. In some cases for chemicals with shallow slope dose-response curves, additional animals (up to a total of fifteen tested) may be needed.

34. When the stopping criteria have been attained, the estimated LD50 should be calculated from the animal outcomes at test termination using the method described in paragraphs 40 and 41.

35. Moribund animals killed for humane reasons are considered in the same way as animals that died on test. If an animal unexpectedly dies late in the study and there are other survivors at that dose or above, it is appropriate to stop dosing and observe all animals to see if other animals will also die during a similar observation period. If subsequent survivors also die, *and* it appears that all dose levels exceed the LD50 it would be most appropriate to start the study again beginning at least two steps below the lowest dose with deaths (and increasing the observation period) since the technique is most accurate when the starting dose is below the LD50. If subseqent animals survive at or above the dose of the animal that dies, it is not necessary to change the dose progression since the information from the animal that has now died will be included into the calculations as a death at a lower dose than subsequent survivors, pulling the LD50 down.

OBSERVATIONS

36. Animals are observed individually at least once during the first 30 minutes after dosing, periodically during the first 24 hours (with special attention given during the first 4 hours), and daily thereafter, for a total of 14 days, except where they need to be removed from the study and humanely killed for animal welfare reasons or are found dead. However, the duration of observation should not be fixed rigidly. It should be determined by the toxic reactions and time of onset and length of recovery period, and may thus be extended when considered necessary. The times at which signs of toxicity appear and disappear are important, especially if there is a tendency for toxic signs to be delayed (17). All observations are systematically recorded with individual records being maintained for each animal.

37. Additional observations will be necessary if the animals continue to display signs of toxicity. Observations should include changes in skin and fur, eyes and mucous membranes, and also respiratory, circulatory, autonomic and central nervous systems, and somatomotor activity and behaviour pattern. Attention should be directed to observations of tremors, convulsions, salivation, diarrhoea, lethargy, sleep and coma. The principles and criteria summarised in the Humane Endpoints Guidance Document (13) should be taken into consideration. Animals found in a moribund condition and animals showing severe pain or enduring signs of severe distress should be humanely killed. When animals are killed for humane reasons or found dead, the time of death should be recorded as precisely as possible.

Body weight

38. Individual weights of animals should be determined shortly before the test substance is administered and at least weekly thereafter. Weight changes should be calculated and recorded. At the end of the test surviving animals are weighed and then humanely killed.

Pathology

39. All animals (including those which die during the test or are removed from the study for animal welfare reasons) should be subjected to gross necropsy. All gross pathological changes should be recorded for each animal. Microscopic examination of organs showing evidence of gross pathology in animals surviving 24 or more hours after the initial dosing may also be considered because it may yield useful information.

DATA AND REPORTING

Data

40. Individual animal data should be provided. Additionally, all data should be summarised in tabular form, showing for each test dose the number of animals used, the number of animals displaying signs of toxicity (17), the number of animals found dead during the test or killed for humane reasons, time of death of individual animals, a description and the time course of toxic effects and reversibility, and necropsy findings. A rationale for the starting dose and the dose progression and any data used to support this choice should be provided.

Calculation of LD50 for the main test

41. The LD50 is calculated using the maximum likelihood method (14)(15), except in the exceptional cases described in paragraph 42. The following statistical details may be helpful in implementing the maximum likelihood calculations suggested (with an assumed *sigma*). All deaths, whether immediate or delayed or humane kills, are incorporated for the purpose of the maximum likelihood analysis. Following Dixon (4), the likelihood function is written as follows:

$L = L_1 L_2 \ldots L_n$,

where

L is the likelihood of the experimental outcome, given *mu* and *sigma*, and n the total number of animals tested.

$L_i = 1 - F(Z_i)$ if the i^{th} animal survived, or
$L_i = F(Z_i)$ if the i^{th} animal died,

where

F = cumulative standard normal distribution,
$Z_i = [\log(d_i) - mu\ ] / sigma$
d_i = dose given to the i^{th} animal, and
sigma = standard deviation in log units of dose (which is not the log standard deviation).

An estimate of the true LD50 is given by the value of *mu* that maximizes the likelihood L (see paragraph 43).

An estimate of *sigma* of 0.5 is used unless a better generic or case-specific value is available.

42. Under some circumstances, statistical computation will not be possible or will likely give erroneous results. Special means to determine/report an estimated LD50 are available for these circumstances as follows:

(a) If testing stopped based on criterion (a) in paragraph 33 (i.e., a boundary dose was tested repeatedly), or if the upper bound dose ended testing, then the LD50 is reported to be above the upper bound. Classification is completed on this basis.

(b) If all the dead animals have higher doses than all the live animals (or if all live animals have higher doses than all the dead animals, although this is practically unlikely), then the LD50 is between the doses for the live and the dead animals. These observations give no further information on the exact value of the LD50. Still, a maximum likelihood LD50 estimate can be

made provided there is a value for *sigma*. Stopping criterion (b) in paragraph 33 describes one such circumstance.

(c) If the live and dead animals have only one dose in common and all the other dead animals have higher doses and all the other live animals lower doses, or vice versa, then the LD50 equals their common dose. If a closely related substance is tested, testing should proceed with a smaller dose progression.

If none of the above situations occurs, then the LD50 is calculated using the maximum likelihood method.

43. Maximum likelihood calculation can be performed using either SAS (14) (e.g., PROC NLIN) or BMDP (15) (e.g., program AR) computer program packages as described in Appendix 1D in Reference 3. Other computer programs may also be used. Typical instructions for these packages are given in appendices to the ASTM Standard E 1163-87 (6). [The *sigma* used in the BASIC program in (6) will need to be edited to reflect the parameters of this OECD 425 Guideline.] The program's output is an estimate of log(LD50) and its standard error.

44. The likelihood-ratio stopping rule (c) in paragraph 33 is based on three measures of test progress, that are of the form of the likelihood in paragraph 41 with different values for *mu*. Comparisons are made after each animal tested after the sixth that does not already satisfy criterion (a) or (b) of paragraph 33. The equations for the likelihood-ratio criteria are provided in Annex 3. These comparisons are most readily performed in an automated manner and can be executed repeatedly, for instance, by a spreadsheet routine such as that also provided in Annex 3. If the criterion is met, testing stops and the LD50 can be calculated by the maximum likelihood method.

Computation of confidence interval

45. Following the main test and estimated LD50 calculation, it may be possible to compute interval estimates for the LD50. Any of these confidence intervals provides valuable information on the reliability and utility of the main test that was conducted. A wide confidence interval indicates that there is more uncertainty associated with the estimated LD50. The reliability of the estimated LD50 is low and the usefulness of the estimated LD50 may be marginal. A narrow interval indicates that there is relatively little uncertainty associated with the estimated LD50. The reliability of the estimated LD50 is high and the usefulness of the estimated LD50 is good. This means that if the main test were to be repeated, the new estimated LD50 should be close to the original estimated LD50 and both of these estimates should be close to the true LD50.

46. Depending on the outcome of the main test, one of two different types of interval estimates of the true LD50 is calculated.

- When at least three different doses have been tested and the middle dose has at least one animal that survived and one animal that died, a profile-likelihood-based computational procedure is used to obtain a confidence interval that is expected to contain the true LD50 95% of the time. However, because small numbers of animals are expected to be used, the actual level of confidence is generally not exact (18). The random stopping rule improves the ability of the test overall to respond to varying underlying conditions, but also causes the reported level of confidence and the actual level of confidence to differ somewhat (19).

- If all animals survive at or below a given dose level and all animals die when dosed at the next higher dose level, an interval is calculated that has as its lower limit the highest dose

tested where all the animals survive and has as its upper limit the dose level where all the animals died. This interval is labeled as "approximate." The exact confidence level associated with this interval cannot be specifically determined. However, because this type of response would only occur when the dose response is steep, in most cases, the true LD50 is expected to be contained within the calculated interval or be very close to it. This interval will be relatively narrow and sufficiently accurate for most practical use.

47. In some instances, confidence intervals are reported as infinite, through including either zero as its lower end or infinity as its upper end, or both. Such intervals, for example, may occur when all animals die or all animals live. Implementing this set of procedures requires specialized computation which is either by use of a dedicated program to be available from the USEPA or OECD or developed following technical details available from the USEPA or OECD (20). Achieved coverage of these intervals and properties of the dedicated program are described in reports (21) also available through the USEPA.

Test report

48. The test report must include the following information:

Test substance:

- physical nature, purity and,where relevant, physico-chemical properties (including isomerisation);
- identification data, including CAS number.

Vehicle (if appropriate):

- justification for choice of vehicle, if other than water.

Test animals:

- species/strain used;
- microbiological status of the animals, when known;
- number, age and sex of animals (including, where appropriate, a rationale for use of males instead of females);
- source, housing conditions, diet, etc.;

Test conditions:

- rationale for initial dose level selection, dose progression factor and for follow-up dose levels
- details of test substance formulation including details of the physical form of the material administered.;
- details of the administration of the test substance including dosing volumes and time of dosing;
- details of food and water quality (including diet type/source, water source).

Results:

- body weight/body weight changes;
- tabulation of response data and dose level for each animal (i.e., animals showing signs of toxicity including nature, severity, duration of effects, and mortality);

- individual weights of animals at the day of dosing, in weekly intervals thereafter, and at the time of death or sacrifice ;
- time course of onset of signs of toxicity and whether these were reversible for each animal;
- necropsy findings and any histopathological findings for each animal, if available;
- LD50 data;
- statistical treatment of results (description of computer routine used and spreadsheet tabulation of calculations).

Discussion and interpretation of results.

Conclusions.

LITERATURE

(1) Dixon W.J. and A.M. Mood. (1948). A Method for Obtaining and Analyzing Sensitivity Data. J. Amer. Statist. Assoc., 43, 109-126.

(2) Dixon W.J. The Up-and-Down Method for Small Samples (1965). J. Amer. Statist. Assoc. 60, 967-978.

(3) Dixon W.J. (1991). Staircase Bioassay: The Up-and-Down Method. Neurosci. Biobehav. Rev., 15, 47-50.

(4) Dixon W.J. (1991) Design and Analysis of Quantal Dose-Response Experiments (with Emphasis on Staircase Designs). Dixon Statistical Associates, Los Angeles CA, USA.

(5) Bruce R.D. (1985). An Up-and-Down Procedure for Acute Toxicity Testing. Fundam. Appl. Tox., 5, 151-157.

(6) ASTM (1987). E 1163-87, Standard Test Method for Estimating Acute Oral Toxicity in Rats. American Society for Testing and Materials, Philadelphia Pa, USA.

(7) Lipnick R.L., J.A. Cotruvo, R.N. Hill, R.D. Bruce, K.A. Stitzel, A.P. Walker, I. Chu, M. Goddard, L. Segal, J.A. Springer, and R.C. Myers. (1995). Comparison of the Up-and-Down, Conventional LD_{50} and Fixed Dose Acute Toxicity Procedures. Fd. Chem. Toxicol., 33, 223-231.

(8) Choi, S.C. (1990). Interval estimation of the LD_{50} based on an up-and-down experiment. Biometrics 46, 485-492.

(9) Vågerö, M. and R. Sundberg. (1999). The distribution of the maximum likelihood estimator in up-and-down experiments for quantal dose-response data. J. Biopharmaceut. Statist. 9(3), 499-519.

(10) Hsi, B.P. (1969). The multiple sample up-and-down method in bioassay. J. Amer. Statist. Assoc. 64, 147-162.

(11) Noordwijk, A.J. van and J. van Noordwijk. (1988). An accurate method for estimating an approximate lethal dose with few animals, tested with a Monte Carlo procedure. Arch. Toxicol. 61, 333-343.

(12) OECD (2000) Guidance Document on Acute Oral Toxicity . Environmental Health and Safety Monograph Series on Testing and Assessment No 24.

(13) OECD (2000) Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation. Environmental Health and Safety Monograph Series on Testing and Assessment No 19.

(14) SAS Institute Inc. (1990). SAS/STAT® User's Guide. Version 6, Fourth Ed. or later. Cary, NC, USA.

(15) BMDP Statistics Software, Inc. (1990). BMDP Statistical Software Manual. W.J. Dixon, Chief Ed. 1990 rev. or later. University of California Press, Berkeley, CA, USA.

(16) OECD(1998) Harmonized Integrated Hazard Classification System for Human Health and Environmentla Effects of Chemical Substances as endorsed by the 28th Joint Meeting of the Chemicals Committee and Working Party on Chemicals in November 1998, Part 2, pg 11. [http://webnet1.oecd.org/oecd/pages/home/displaygeneral/0,3380,EN-documents-521-14-no-24-no-0,FF.html].

(17) Chan P.K. and A.W. Hayes. (1994). Chap. 16. Acute Toxicity and Eye Irritancy. *Principles and Methods of Toxicology*. Third Edition. A.W. Hayes, Editor. Raven Press, Ltd., New York, USA.

(18) Rosenberger, W.F., Flournoy, N. and Durham, S.D. (1997). Asymptotic normality of maximum likelihood estimators from multiparameter response-driven designs. Journal of Statistical Planning and Inference 60, 69-76.

(19) Jennison, C. and B.W. Turnbull. 2000. Group Sequential Methods with Application to Clinical Trials. Chapman & Hall/CRC: Boca Raton, FL. USA.

(20) Acute Oral Toxicity (OECD Test Guideline 425) Statistical Programme (AOT 425 StatPgm). Version: 1.0, 2001. [http://www.oecd.org/oecd/pages/home/displaygeneral/0,3380,EN-document-524-nodirectorate-no-24-6775-8,FF.html]

(21) Westat. 2001. Simulation Results from the AOT425StatPgm Program. Report prepared for U.S. E.P.A. under Contract 68-W7-0025, Task Order 5-03.

OECD/OCDE **425**

Adopted:
17th December 2001

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.

Delayed death means that an animal does not die or appear moribund within 48 hours but dies later during the 14-day observation period.

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).

Dose progression factor, sometimes termed a dose spacing factor, refers to the multiple by which a dose is increased (i.e., the dose progression) when an animal survives or the divisor by which it is decreased when an animal dies. The dose progression factor is recommended to be the antilog of 1/ (the estimated slope of the dose response curve). The default dose progression factor is recommended to be 3.2 = antilog 0.5 = antilog ½.

GHS: Globally Harmonised Classification System for Chemical Substances and Mixtures. A joint activity of OECD (human health and the environment), UN Committee of Experts on Transport of Dangerous Goods (physical–chemical properties) and ILO (hazard communication) and co-ordinated by the Interorganisation Programme for the Sound Management of Chemicals (IOMC).

Impending death: when moribund state or death is expected prior to the next planned time of observation. Signs indicative of this state in rodents could include convulsions, lateral position, recumbence, and tremor. (See the Humane Endpoint Guidance Document (13) for more details).

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 (2000 or 5000 mg/kg).

Moribund status : being in a state of dying or inability to survive, even if treated. (See the Humane Endpoint Guidance Document (13) for more details).

Nominal sample size refers to the total number of tested animals, reduced by one less than the number of like responses at the beginning of the series, or by the number of tested animals up to but not including the pair that creates the first reversal. For example, for a series where X and O indicate opposite animal outcomes (for instance, X could be: "dies within 48 hours" and O: " survives") in a pattern as follows: OOOXXXOXO, we have the total number of tested animals (or sample size in the conventional sense) as 8 and the nominal sample size as 6. This particular example shows 4 animals following a reversal. It is important to note whether a count in a particular part of the guideline refers to the nominal sample size or to the total number tested. For example, the maximum actual number tested is 15. When testing is stopped based on that maximum number, the nominal sample size will be less than or equal to 15. Members of the nominal sample start with the (r-1)st animal (the animal before the second in the reversal pair) (see reversal below).

Predictable death: presence of clinical signs indicative of death at a known time in the future before the planned end of the experiment, for example: inability to reach water or food. (See the Humane Endpoint Guidance Document (13) for more details).

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.

Reversal is a situation where nonresponse is observed at some dose, and a response is observed at the next dose tested, or vice versa (i.e., response followed by nonresponse). Thus, a reversal is created by a pair of responses. The first such pair occurs at animals numbered r-1 and r.

Sigma is the standard deviation of a log normal curve describing the range of tolerances of test subjects to the chemical (where a subject is expected capable of responding if the chemical dose exceeds the subject's tolerance). The estimated *sigma* provides an estimate of the variation among test animals in response to a full range of doses.
See slope and probit.

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.*

Stopping rule is used in this guideline synonymously with 1) a specific stopping criterion and 2) the collection of all criteria determining when a testing sequence terminates. In particular, for the main test, stopping rule is used in paragraph 7 as a shorthand for the criterion that relies on comparison of ratios to a critical value.

OECD/OCDE 425

ANNEX 2

DOSING PROCEDURE

Dose Sequence for Main Test

1. Up-and-Down Dosing Procedure. For each run, animals are dosed, one at a time, usually at 48-hour intervals. The first animal receives a dose a step below the level of the best estimate of the LD50. This selection reflects an adjustment for a tendency to bias **away from the LD50 in the direction of** the initial starting dose in the final estimate (see paragraph 7 of the Guideline). The overall pattern of outcomes is expected to stabilize as dosing is adjusted for each subsequent animal. Paragraph 3 below provides further guidance for choice of dose spacing factor.

2. Default Dose Progression. Once the starting dose and dose spacing are decided, the toxicologist should list all possible doses including the upper bound (usually 2000 or 5000 mg/kg). Doses that are close to the upper bound should be removed from the progression. The stepped nature of the TG 425 design provides for the first few doses to function as a self-adjusting sequence. Because of the tendency for positive bias, in the event that nothing is known about the substance, a starting dose of 175 mg/kg is recommended. If the default procedure is to be used for the main test, dosing will be initiated at 175 mg/kg and doses will be spaced by a factor of 0.5 on a log dose scale. The doses to be used include 1.75, 5.5, 17.5, 55, 175, 550, 2000 or, for specific regulatory needs, 1.75, 5.5, 17.5, 55, 175, 550, 1750, 5000. For certain highly toxic substances, the dosing sequence may need to be extended to lower values.

3. In the event a dose progression factor other than the default is deemed suitable, Table 1 provides dose progressions for whole number multiples of slope, from 1 to 8.

425 OECD/OCDE

Table 1 Dose Progressions for OECD Guideline 425
Choose a Slope and Read Down the Column
All doses in mg/kg bw

Slope = 1	2	3	4	5	6	7	8
0.175*	0.175*	0.175*	0.175*	0.175*	0.175*	0.175*	0.175*
						0.24	0.23
				0.275	0.26		
			0.31			0.34	0.31
		0.375			0.375		
							0.41
				0.44		0.47	
	0.55		0.55		0.55		0.55
				0.69		0.65	
							0.73
		0.81			0.82		
			0.99			0.91	0.97
				1.09	1.2		
						1.26	1.29
1.75	1.75	1.75	1.75	1.75	1.75	1.75	1.75
						2.4	2.3
				2.75	2.6		
			3.1			3.4	3.1
		3.75			3.75		
				4.4			4.1
						4.7	
	5.5		5.5		5.5		5.5
				6.9		6.5	
							7.3
		8.1			8.2		
			9.9			9.1	9.7
				10.9	12		
						12.6	12.9
17.5	17.5	17.5	17.5	17.5	17.5	17.5	17.5
						24	23
				27.5	26		
			31			34	31

OECD/OCDE **425**

Table 1 continued

		37.5			37.5		
				44			41
						47	
	55		55		55		55
						65	
				69			73
		81			82		
			99			91	97
				109	120		
						126	129
175	175	175	175	175	175	175	175
						240	230
				275	260		
			310			340	310
		375			375		
				440			410
						470	
	550		550		550		550
						650	
				690			730
		810			820		
			990			910	970
				1090	1200		
						1260	1290
1750	1750	1750	1750	1750	1750	1750	1750
						2400	2300
				2750	2600		
			3100				3100
					3750	3400	
							4100
5000	5000	5000	5000	5000	5000	5000	5000

* If lower doses are needed, continue progressions to a lower dose

425 OECD/OCDE

ANNEX 3

COMPUTATIONS FOR THE LIKELIHOOD-RATIO STOPPING RULE

1. As described in Guideline paragraph 33, the main test may be completed on the basis of the first of three stopping criteria to occur. In any case, even if none of the stopping criteria is satisfied, dosing would stop when 15 animals are dosed. Tables 2-5 illustrate examples where testing has started with no information, so the recommended default starting value, 175 mg/kg, and the recommended default dose progression factor, 3.2 or one half log, have been used. Please note the formatting of these tables is only illustrative.

2. Table 2 shows how the main test would stop if 3 animals have survived at the limit dose of 2000 mg/kg; Table 3 shows a similar situation when the limit dose of 5000 mg/kg is used. (These illustrate situations where a Limit Test was not thought appropriate *a priori*.) Table 4 shows how a particular sequence of 5 reversals in 6 tested animals could occur and allow test completion. Finally, Table 5 illustrates a situation where neither criterion (a) nor criterion (b) has been met, a reversal of response has occurred followed by 4 tested animals, and, consequently, criterion (c) must be evaluated as well.

3. Criterion (c) calls for a likelihood-ratio stopping rule to be evaluated after testing each animal, starting with the fourth tested following the reversal. Three "measures of test progress" are calculated. Technically, these measures of progress are likelihoods, as recommended for the maximum-likelihood estimation of the LD50. The procedure is closely related to calculation of a confidence interval by a likelihood-based procedure.

4. The basis of the procedure is that when enough data have been collected, a point estimate of the LD50 should be more strongly supported than values above and below the point estimate, where statistical support is quantified using likelihood. Therefore three likelihood values are calculated: a likelihood at an LD50 point estimate (called the rough estimate or dose-averaging estimate in the example), a likelihood at a value below the point estimate, and a likelihood at a value above the point estimate. Specifically, the low value is taken to be the point estimate divided by 2.5 and the high value is taken to be the point estimate multiplied by 2.5.

5. The likelihood values are compared by calculating ratios of likelihoods, and then determining whether these likelihood-ratios (LR) exceed a critical value. Testing stops when the ratio of the likelihood for the point estimate exceeds each of the other likelihoods by a factor of 2.5, which is taken to indicate relatively strong statistical support for the point estimate. Therefore two likelihood-ratios (LRs) are calculated, a ratio of likelihoods for the point estimate and the point estimate divided by 2.5, and a ratio for the point estimate and the estimate times 2.5.

6. The calculations are easily performed in any spreadsheet with normal probability functions. The calculations are illustrated in Table 5, which is structured to promote spreadsheet implementation. The computation steps are illustrated using an example where the upper limit dose is 5000 mg/kg, but the computational steps are carried out in the same fashion when the upper boundary dose is 2000 mg/kg. Empty spreadsheets preprogrammed with the necessary formulas are available for direct downloading on the OECD and EPA web sites.

Hypothetical example using an upper limit dose of 5000 mg/kg (Table 5)

7. In the hypothetical example utilizing an upper boundary dose of 5000 mg/kg, the LR stopping criterion was met after nine animals had been tested. The first "reversal" occurred with the 3rd animal tested. The LR stopping criterion is checked when four animals have been tested following the reversal. In this example, the fourth animal tested following the reversal is the seventh animal actually tested. Therefore, for this example, the spreadsheet calculations are only needed after the seventh animal had been tested and the data could be entered at that time. Subsequently, the LR stopping criterion would have been checked after testing the seventh animal, the eighth animal, and the ninth. The LR stopping criterion is first satisfied after the ninth animal is tested in this example.

A. Enter the dose-response information animal by animal.

Column 1. Steps are numbered 1-15. No more than 15 animals may be tested.
Column 2. Place an I in this column as each animal is tested.
Column 3. Enter the dose received by the i^{th} animal.
Column 4. Indicate whether the animal responded (shown by an X) or did not respond (shown by an O).

B. The nominal and actual sample sizes.

8. The nominal sample consists of the two animals that represent the first reversal (here the second and third animals), plus all animals tested subsequently. Here, Column 5 indicates whether or not a given animal is included in the nominal sample.

The nominal sample size (nominal n) appears in Row 16. This is the number of animals in the nominal sample. In the example, nominal n is 8.
The actual number tested appears in Row 17.

C. Rough estimate of the LD50.

9. The geometric mean of doses for the animals in the current nominal sample is used as a rough estimate of the LD50 from which to gauge progress. In the table, this is called the "dose-averaging estimator." It is updated with each animal tested. This average is restricted to the nominal sample in order to allow for a poor choice of initial test dose, which could generate either an initial string of responses or an initial string of nonresponses. (However, the results for all animals are used in the likelihood calculations for final LD50 calculation below.) Recall that the geometric mean of n numbers is the product of the n numbers, raised to a power of $1/n$.

The dose-averaging estimate appears in Row 18 (e.g., $(175 * 550 * ... * 1750)^{1/8} = 1292.78$).
Row 19 shows the logarithm (base 10) of the value in Row 18 (e.g., $\log_{10} 1292.8 = 3.112$).

D. Likelihood for the rough LD50 estimate.

10. Likelihood is a statistical measure of how strongly the data support an estimate of the LD50 or other parameter. Ratios of likelihood values can be used to compare how well the data support different estimates of the LD50.

11. In column 8 calculate the likelihood for Step C's rough LD50 estimate. The likelihood (Row 21) is the product of likelihood contributions for individual animals (see Guideline paragraph 41). The likelihood contribution for the i^{th} animal is denoted L_i.

12. In column 7 enter the estimate of the probability of response at dose d_i, denoted P_i. P_i is calculated from a dose-response curve. Note that the parameters of a probit dose-response curve are the

slope and the LD50, so values are needed for each of those parameters. For the LD50 the dose-averaging estimate from Row 18 is used. For the slope in this example the default value of 2 is used. The following steps may be used to calculate the response probability P_i.

1. Calculate the base-10 log of dose d_i (Column 6).

2. For each animal calculate the z-score, denoted Z_i (not shown in the table), using the formulae
   $sigma = 1 / slope$,
   $Z_i = (\log_{10}(d_i) - \log_{10}(LD50)) / sigma$

   For example, for the first animal (Row 1),
   $sigma = 1 / 2$
   $Z_1 = (2.243 - 3.112) / 0.500 = -1.738$

3. For the i^{th} dose the estimated response probability is

   $P_i = F(Z_i)$

where F denotes the cumulative distribution function for the standard normal distribution (i.e., the normal distribution with mean 0 and variance 1).

For example (Row 1),

$P_1 = F(-1.738) = 0.0412$

The function F (or something very close) is ordinarily what is given for the normal distribution in statistical tables, but the function is also widely available as a spreadsheet function. It is available under different names, for example the @NORMAL function of Lotus 1-2-3 (1) and the @NORMDIST function in Excel (2). To confirm that you have used correctly the function available in your software, you may wish to verify familiar values such as $F(1.96) \approx 0.975$ or $F(1.64) \approx 0.95$.

13. Column 8. Calculate the natural log of the likelihood contribution ($\ln(L_i)$). L_i is simply the probability of the response that actually was observed for the i^{th} animal:

responding animals: $\ln(L_i) = \ln (P_i)$
non-responding animals: $\ln(L_i) = \ln(1 - P_i)$

Note that here the natural logarithm (ln) is used, whereas elsewhere the base-10 (common) logarithm was used. These choices are what are ordinarily expected in a given context.

The steps above are performed for each animal. Finally:

Row 20: Sum the log-likelihood contributions in Column 8.
Row 21: Calculate the likelihood by applying the exp function applied to the log-likelihood value in Row 20 (e.g., $\exp(-3.389) = e^{-3.389} = 0.0337$).

E. Calculate likelihoods for two dose values above and below the rough estimate.

14. If the data permit a precise estimate, then one expects the likelihood should be high if the estimate is a reasonable estimate of the LD50, relative to likelihoods for values distant from this estimate. Compare the likelihood for the dose-averaging estimate (1292.8, Row 18) to values differing by a factor of

2.5 from that value (i.e., to 1292.8*2.5 and 1292.8/2.5). The calculations (displayed in Columns 9-12) are carried out in a fashion similar to those described above, except that the values 517.1 (=1292.8/2.5) and 3232.0 (=1292.8*2.5) have been used for the LD50, instead of 1292.8. The likelihoods and log-likelihoods are displayed in Rows 20-21.

F. Calculate likelihood-ratios.

15. The three likelihood values (Row 21) are used to calculate two likelihood-ratios (Row 22). A likelihood-ratio is used to compare the statistical support for the estimate of 1292.8 to the support for each of the other values, 517.1 and 3232.0. The two likelihood-ratios are therefore:

LR1 = [likelihood of 1292.8] / [likelihood of 517.1]
= 0.0337 / 0.0080
= 4.21

and

LR2 = [likelihood of 1292.8] / [likelihood of 3232.0]
= 0.0337 / 0.0098
= 3.44

G. Determine if the likelihood-ratios exceed the critical value.

16. High likelihood-ratios are taken to indicate relatively high support for the point estimate of the LD50. Both of the likelihood-ratios calculated in Step F (4.21 and 3.44) exceed the critical likelihood-ratio, which is 2.5. Therefore the LR stopping criterion is satisfied and testing stops. This is indicated by a TRUE in Row 24 and a note at the top of the example spreadsheet that the LR criterion is met.

LITERATURE

(1) Lotus Development Corporation (1999). Lotus® 1-2-3. Version 9.5, Millenium Edition. Cambridge, MA, USA.

(2) Microsoft Corporation (1985-1997). Microsoft® Excel Version 5.0 or later. Seattle, WA, USA.

Table 2. Example of stopping criterion (a) using 2000 mg/kg.

Stop after animal #5 because 3 animals survive at limit of 2000 mg/kg (#3-#5).

1	2	3	4	5	6	7	8	9	10	11	12
Step	(I)nclude; (E)xclude	Dose	(X)response (O)non-resp.	Included in nominal *n*	log10 Dose	LD50 = Prob. of response	#DIV/0! likelihood contribn. (ln *Li*)	LD50 = Prob. of response	#DIV/0! likelihood contribn. (ln *Li*)	LD50 = Prob. of response	#DIV/0! likelihood contribn. (ln *Li*)
			OK								
1	I	175	O	no	2.2430	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
2	I	550	O	no	2.7404	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
3	I	2000	O	no	3.3010	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
4	I	2000	O	no	3.3010	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
5	I	2000	O	no	3.3010	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
6	E				-						
7	E				-						
8	E				-	-	-	-	-	-	-
9	E				-	-	-	-	-	-	-
10	E				-	-	-	-	-	-	-
11	E				-	-	-	-	-	-	-
12	E				-	-	-	-	-	-	-
13	E				-	-	-	-	-	-	-
14	E				-	-	-	-	-	-	-
15	E				-	-	-	-	-	-	-
Nominal Sample size =				0							
Actual number tested =				5							

Ignore all calculation cells. No reversal in direction of response.

Maximum Likelihood Calculations cannot be completed. LD50 is greater than 2000 mg/kg.

Calculated maximum likelihood estimate of LD50 = none

Table 3. Example of stopping criterion (a) using 5000 mg/kg.

Stop after animal #6 because 3 animals survive at limit of 5000 mg/kg (#4-#6).

1	2	3	4	5	6	7	8	9	10	11	12
Step	(I)nclude; (E)xclude	Dose	(X)response (O)non-resp. OK	Included in nominal *n*	log10 Dose	LD50 = Prob. of response	#DIV/0! likelihood contribn. (ln *Li*)	LD50 = Prob. of response	#DIV/0! likelihood contribn. (ln *Li*)	LD50 = Prob. of response	#DIV/0! likelihood contribn. (ln *Li*)
1	I	175	O	no	2.2430	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
2	I	550	O	no	2.7404	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
3	I	1750	O	no	3.2430	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
4	I	5000	O	no	3.6990	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
5	I	5000	O	no	3.6990	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
6	I	5000	O	no	3.6990	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
7	E				-	Ignore all calculation cells. No reversal in direction of response.					
8	E				-						
9	E				-	-	-	-	-	-	-
10	E				-	-	-	-	-	-	-
11	E				-	-	-	-	-	-	-
12	E				-	Maximum Likelihood Calculations			-	-	-
13	E				-	cannot be completed. LD50 is			-	-	-
14	E				-	greater than 5000 mg/kg.			-	-	-
15	E				-	-	-	-	-	-	-
Nominal Sample size =				0							
Actual number tested =				6							

Calculated maximum likelihood estimate of LD50 = none

Table 4. Example of stopping criterion (b)

Stop after animal #7 because 5 reversals in 6 consecutive animals tested (#2-#7).

1	2	3	4	5	6	7	8	9	10	11	12
Step	(I)nclude; (E)xclude	Dose	(X)response (O)non-resp. OK	Included in nominal *n*	log10 Dose	LD50 = Prob. of response	31.0 likelihood contribn. (ln *Li*)	LD50 = Prob. of response	12.4 likelihood contribn. (ln *Li*)	LD50 = Prob. of response	77.6 likelihood contribn. (ln *Li*)
1	I	175	X	no	2.2430	0.9335	-0.0688	0.9892	-0.0108	0.7602	-0.2742
2	I	55	X	yes	1.7404	0.6905	-0.3703	0.9020	-0.1031	0.3826	-0.9607
3	I	17.5	O	yes	1.2430	0.3095	-0.3703	0.6174	-0.9607	0.0980	-0.1031
4	I	55	X	yes	1.7404	0.6905	-0.3703	0.9020	-0.1031	0.3826	-0.9607
5	I	17.5	O	yes	1.2430	0.3095	-0.3703	0.6174	-0.9607	0.0980	-0.1031
6	I	55	X	yes	1.7404	0.6905	-0.3703	0.9020	-0.1031	0.3826	-0.9607
7	I	17.5	O	yes	1.2430	0.3095	-0.3703	0.6174	-0.9607	0.0980	-0.1031
8	E				-	-	-	-	-	-	-
9	E				-	-	-	-	-	-	-
10	E				-	-	-	-	-	-	-
11	E				-	-	-	-	-	-	-
12	E				-	-	-	-	-	-	-
13	E				-	-	-	-	-	-	-
14	E				-	-	-	-	-	-	-
15	E				-	-	-	-	-	-	-
Nominal Sample size =				6							
Actual number tested =				7							
Dose-averaging estimator				31.02							
log10 =				1.492							
log-likelihood sums:							-2.2906		-3.2021		-3.4655
likelihoods:							0.1012		0.0407		0.0313
likelihood ratios:									2.4880		3.2378
Individual ratios exceed critical value?				critical=	2.5		Automated calculation; not relevant to this case.		FALSE		TRUE
Both ratios exceed critical value?									FALSE		
Calculated maximum likelihood estimate of LD50 =					29.6	Final estimate obtained from Maximum Likelihood Calculations					

Table 5. Example of stopping criterion (c)

Stop when LR criterion is first met, here at animal #9.
Check LR criterion starting at animal #6.

Assumed slope	2	sigma =	0.5

Parameters of convergence criterion	
critical LR	2.5
factor of LD50	2.5

Result: The LR criterion is met

1	2	3	4	5	6		7	8	9	10	11	12
Step	(I)nclude; (E)xclude	Dose	(X)response (O)non-resp. OK	Included in nominal *n*	log10 Dose	Contrib.to DAE	LD50 = Prob. of response	1292.8 likelihood contribn. (ln *Li*)	LD50 = Prob. of response	517.1 likelihood contribn. (ln *Li*)	LD50 = Prob. of response	3232.0 likelihood contribn. (ln *Li*)
1	I	175	O	no	2.2430	0.0000	0.0412	-0.0421	0.1733	-0.1903	0.0057	-0.0057
2	I	550	O	yes	2.7404	2.7404	0.2289	-0.2600	0.5214	-0.7368	0.0620	-0.0640
3	I	1750	X	yes	3.2430	3.2430	0.6037	-0.5046	0.8552	-0.1564	0.2971	-1.2138
4	I	550	O	yes	2.7404	2.7404	0.2289	-0.2600	0.5214	-0.7368	0.0620	-0.0640
5	I	1750	X	yes	3.2430	3.2430	0.6037	-0.5046	0.8552	-0.1564	0.2971	-1.2138
6	I	550	O	yes	2.7404	2.7404	0.2289	-0.2600	0.5214	-0.7368	0.0620	-0.0640
7	I	1750	O	yes	3.2430	3.2430	0.6037	-0.9257	0.8552	-1.9323	0.2971	-0.3525
8	I	5000	X	yes	3.6990	3.6990	0.8800	-0.1279	0.9756	-0.0247	0.6477	-0.4344
9	I	1750	X	yes	3.2430	3.2430	0.6037	-0.5046	0.8552	-0.1564	0.2971	-1.2138
10	E				-	0.0000	-	-	-	-	-	-
11	E				-	0.0000	-	-	-	-	-	-
12	E				-	0.0000	-	-	-	-	-	-
13	E				-	0.0000	-	-	-	-	-	-
14	E				-	0.0000	-	-	-	-	-	-
15	E				-	0.0000	-	-	-	-	-	-
Nominal Sample size =				8								
Actual number tested =				9								
Dose-averaging estimator				1292.78								
log10 =				3.112								
log-likelihood sums:								-3.3894		-4.8270		-4.6260
likelihoods:								0.0337		0.0080		0.0098
likelihood ratios:										4.2104		3.4436
Individual ratios exceed critical value?				critical=	2.5					TRUE		TRUE
Both ratios exceed critical value?										TRUE		

425 **OECD/OCDE**

ANNEX 4

CRITERIA FOR CLASSIFICATION OF TEST SUBSTANCES WITH EXPECTED LD50 VALUES EXCEEDING 2000 MG/KG WITHOUT THE NEED FOR TESTING

1. Criteria for hazard Category 5 are intended to enable the identification of test substances which are of relatively low acute toxicity hazard but which, under certain circumstances may present a danger to vulnerable populations. These substances are anticipated to have an oral or dermal LD50 in the range of 2000-5000 mg/kg or equivalent doses for other routes. Test substances could be classified in the hazard category defined by: 2000 mg/kg<LD50<5000 mg/kg (Category 5 in the GHS) in the following cases:

a) if reliable evidence is already available that indicates the LD50 to be in the range of Category 5 values; or other animal studies or toxic effects in humans indicate a concern for human health of an acute nature.

b) through extrapolation, estimation or measurement of data if assignment to a more hazardous category is not warranted, and
   - reliable information is available indicating significant toxic effects in humans, or
   - any mortality is observed when tested up to Category 4 values by the oral route, or
   - where expert judgement confirms significant clinical signs of toxicity, when tested up to Category 4 values, except for diarrhoea, piloerection or an ungroomed appearance, or
   - where expert judgement confirms reliable information indicating the potential for significant acute effect from the other animal studies.

TESTING AT DOSES ABOVE 2000 MG/KG

2. Recognising the need to protect animal welfare, testing in animals in Category 5 ranges is discouraged and should only be considered when there is a strong likelihood that results of such a test would have a direct relevance for protecting human health.

APPENDIX B3

OECD GUIDELINE 423: ACUTE ORAL TOXICITY – ACUTE TOXIC CLASS METHOD

[This Page Intentionally Left Blank]

OECD/OCDE **423**

Adopted:
17th December 2001

OECD GUIDELINE FOR TESTING OF CHEMICALS

Acute Oral Toxicity – Acute Toxic Class Method

INTRODUCTION

1. OECD Guidelines for the Testing of Chemicals are periodically reviewed in the light of scientific progress or changing assessment practices. The original Guideline 423 was adopted in March 1996 as the second alternative to the conventional acute toxicity test, described in Test Guideline 401. Based on the recommendations of several expert meetings, revision was considered timely because: i) international agreement has been reached on harmonised LD50 cut-off values for the classification of chemical substances, which differ from the cut-offs recommended in the 1996 version of the Guideline, and ii) testing in one sex (usually females) is now considered sufficient.

2. The acute toxic class method (1) set out in this Guideline is a stepwise procedure with the use of 3 animals of a single sex per step. Depending on the mortality and/or the moribund status of the animals, on average 2-4 steps may be necessary to allow judgement on the acute toxicity of the test substance. This procedure is reproducible, uses very few animals and is able to rank substances in a similar manner to the other acute toxicity testing methods (Test Guidelines 420 and 425). The acute toxic class method is based on biometric evaluations (2)(3)(4)(5) with fixed doses, adequately separated to enable a substance to be ranked for classification purposes and hazard assessment. The method as adopted in 1996 was extensively validated *in vivo* against LD50 data obtained from the literature, both nationally (6) and internationally (7).

3. Guidance on the selection of the most appropriate test method for a given purpose can be found in the Guidance Document on Acute Oral Toxicity Testing (8). This Guidance Document also contains additional information on the conduct and interpretation of Test Guideline 423.

4. Definitions used in the context of this Guideline are set out in Annex 1.

INITIAL CONSIDERATIONS

5. Test substances, at doses that are known to cause marked pain and distress due to corrosive or severely irritant actions, need not be administered. Moribund animals, or animals obviously in pain or showing signs of severe and enduring distress shall be humanely killed, and are considered in the interpretation of the test results in the same way as animals that died on test. Criteria for making the decision to kill moribund or severely suffering animals, and guidance on the recognition of predictable or impending death, are the subject of a separate Guidance Document (9).

6. The method uses pre-defined doses and the results allow a substance to be ranked and classified according to the Globally Harmonised System for the classification of chemicals which cause acute toxicity (10).

7. In principle, the method is not intended to allow the calculation of a precise LD_{50}, but does allow for the determination of defined exposure ranges where lethality is expected since death of a proportion of the animals is still the major endpoint of this test. The method allows for the determination of an LD50 value only when at least two doses result in mortality higher than 0% and lower than 100%. The use of a selection of pre-defined doses, regardless of test substance, with classification explicitly tied to number of animals observed in different states improves the opportunity for laboratory to laboratory reporting consistency and repeatability.

8. The testing laboratory should consider all available information on the test substance prior to conducting the study. Such information will include the identity and chemical structure of the substance; its physico-chemical properties; the result of any other *in vivo* or *in vitro* toxicity tests on the substance; toxicological data on the structurally related substances; and the anticipated use(s) of the substance. This information is necessary to satisfy all concerned that the test is relevant for the protection of human health and will help in the selection of the most appropriate starting dose.

PRINCIPLE OF THE TEST

9. It is the principle of the test that, based on a stepwise procedure with the use of a minimum number of animals per step, sufficient information is obtained on the acute toxicity of the test substance to enable its classification. The substance is administered orally to a group of experimental animals at one of the defined doses. The substance is tested using a stepwise procedure, each step using three animals of a single sex (normally females). Absence or presence of compound-related mortality of the animals dosed at one step will determine the next step, i.e.;

- no further testing is needed,
- dosing of three additional animals, with the same dose
- dosing of three additional animals at the next higher or the next lower dose level.

10. Details of the test procedure are described in Annex 2. The method will enable a judgement with respect to classifying the test substance to one of a series of toxicity classes defined by fixed LD50 cut-off values.

DESCRIPTION OF THE METHOD

Selection of animal species

11. The preferred rodent species is the rat, although other rodent species may be used. Normally females are used (9). This is because literature surveys of conventional LD50 tests show that, although there is little difference in sensitivity between the sexes, in those cases where differences are observed females are generally slightly more sensitive (11). However if knowledge of the toxicological or toxicokinetic properties of structurally related chemicals indicates that males are likely to be more sensitive, then this sex should be used. When the test is conducted in males adequate justification should be provided.

12. Healthy young adult animals of commonly used laboratory strains should be employed. Females should be nulliparous and non-pregnant. Each animal, at the commencement of its dosing, should be between 8 and 12 weeks old and its weight should fall in an interval within $\pm$ 20 % of the mean weight of any previously dosed animals.

Housing and feeding conditions

13. The temperature in the experimental animal room should be 22°C (± 3°C). Although the relative humidity should be at least 30% and preferably not exceed 70% other than during room cleaning the aim should be 50-60%. Lighting should be artificial, the sequence being 12 hours light, 12 hours dark. For feeding, conventional laboratory diets may be used with an unlimited supply of drinking water. Animals may be group-caged by dose, but the number of animals per cage must not interfere with clear observations of each animal.

Preparation of animals

14. The animals are randomly selected, marked to permit individual identification, and kept in their cages for at least 5 days prior to dosing to allow for acclimatisation to the laboratory conditions.

Preparation of doses

15. In general test substances should be administered in a constant volume over the range of doses to be tested by varying the concentration of the dosing preparation. Where a liquid end product or mixture is to be tested however, the use of the undiluted test substance, ie at a constant concentration, may be more relevant to the subsequent risk assessment of that substance, and is a requirement of some regulatory authorities. In either case, the maximum dose volume for administration must not be exceeded. The maximum volume of liquid that can be administered at one time depends on the size of the test animal. In rodents, the volume should not normally exceed 1mL/100g of body weight: however in the case of aqueous solutions 2 mL/100g body weight can be considered. With respect to the formulation of the dosing preparation, the use of an aqueous solution/suspension/emulsion is recommended wherever possible, followed in order of preference by a solution/suspension/emulsion in oil (e.g. corn oil) and then possibly solution in other vehicles. For vehicles other than water the toxicological characteristics of the vehicle should be known. Doses must be prepared shortly prior to administration unless the stability of the preparation over the period during which it will be used is known and shown to be acceptable.

PROCEDURE

Administration of doses.

16. The test substance is administered in a single dose by gavage using a stomach tube or a suitable intubation canula. In the unusual circumstance that a single dose is not possible, the dose may be given in smaller fractions over a period not exceeding 24 hours.

17. Animals should be fasted prior to dosing (e.g. with the rat, food but not water should be withheld over-night, with the mouse, food but not water should be withheld for 3-4 hours). Following the period of fasting, the animals should be weighed and the test substance administered. After the substance has been administered, food may be withheld for a further 3-4 hours in rats or 1-2 hours in mice. Where a dose is administered in fractions over a period it may be necessary to provide the animals with food and water depending on the length of the period.

Number of animals and dose levels

18. Three animals are used for each step. The dose level to be used as the starting dose is selected from one of four fixed levels, 5, 50, 300 and 2000 mg/kg body weight. The starting dose level should be that which is most likely to produce mortality in some of the dosed animals. The flow charts of Annex 2 describe the procedure that should be followed for each of the starting doses.

19. When available information suggests that mortality is unlikely at the highest starting dose level (2000 mg/kg body weight), then a limit test should be conducted. When there is no information on a substance to be tested, for animal welfare reasons it is recommended to use the starting dose of 300 mg/kg body weight.

20 The time interval between treatment groups is determined by the onset, duration, and severity of toxic signs. Treatment of animals at the next dose, should be delayed until one is confident of survival of the previously dosed animals.

21. Exceptionally, and only when justified by specific regulatory needs, the use of additional upper dose level of 5000 mg/kg body weight may be considered (see Annex 3). For reasons of animal welfare concern, testing of animals in GHS Category 5 ranges (2000-5000mg/kg) is discouraged and should only be considered when there is a strong likelihood that results of such a test have a direct relevance for protecting human or animal health or the environment.

Limit test

22. The limit test is primarily used in situations where the experimenter has information indicating that the test material is likely to be nontoxic, i.e., having toxicity only above regulatory limit doses. Information about the toxicity of the test material can be gained from knowledge about similar tested compounds or similar tested mixtures or products, taking into consideration the identity and percentage of components known to be of toxicological significance. In those situations where there is little or no information about its toxicity, or in which the the test material is expected to be toxic, the main test should be performed.

23. A limit test at one dose level of 2000 mg/kg body weight may be carried out with six animals (three animals per step). Exceptionally a limit test at one dose level of 5000 mg/kg may be carried out with three animals (see Annex 3). If test substance-related mortality is produced, further testing at the next lower level may need to be carried out.

OBSERVATIONS

24. Animals are observed individually after dosing at least once during the first 30 minutes, periodically during the first 24 hours, with special attention given during the first 4 hours, and daily thereafter, for a total of 14 days, except where they need to be removed from the study and humanely killed for animal welfare reasons or are found dead. However, the duration of observation should not be fixed rigidly. It should be determined by the toxic reactions, time of onset and length of recovery period, and may thus be extended when considered necessary. The times at which signs of toxicity appear and disappear are important, especially if there is a tendency for toxic signs to be delayed (12). All observations are systematically recorded with individual records being maintained for each animal.

25. Additional observations will be necessary if the animals continue to display signs of toxicity. Observations should include changes in skin and fur, eyes and mucous membranes, and also respiratory, circulatory, autonomic and central nervous systems, and somatomotor activity and behaviour pattern. Attention should be directed to observations of tremors, convulsions, salivation, diarrhoea, lethargy, sleep

and coma. The principles and criteria summarised in the Humane Endpoints Guidance Document (9) should be taken into consideration. Animals found in a moribund condition and animals showing severe pain or enduring signs of severe distress should be humanely killed. When animals are killed for humane reasons or found dead, the time of death should be recorded as precisely as possible.

Body weight

26. Individual weights of animals should be determined shortly before the test substance is administered, and at least weekly thereafter. Weight changes should be calculated and recorded. At the end of the test surviving animals are weighed and humanely killed.

Pathology

27. All test animals (including those that die during the test or are removed from the study for animal welfare reasons) should be subjected to gross necropsy. All gross pathological changes should be recorded for each animal. Microscopic examination of organs showing evidence of gross pathology in animals surviving 24 or more hours may also be considered because it may yield useful information.

DATA AND REPORTING

Data

28. Individual animal data should be provided. Additionally, all data should be summarised in tabular form, showing for each test group the number of animals used, the number of animals displaying signs of toxicity, the number of animals found dead during the test or killed for humane reasons, time of death of individual animals, a description and the time course of toxic effects and reversibility, and necropsy findings.

Test report

29. The test report must include the following information, as appropriate:

Test substance:

- physical nature, purity, and, where relevant, physico-chemical properties (including isomerisation);
- identification data, including CAS number.

Vehicle (if appropriate):

- justification for choice of vehicle, if other than water.

Test animals:

- species/strain used;
- microbiological status of the animals, when known;
- number, age, and sex of animals (including, where appropriate, a rationale for the use of males instead of females);
- source, housing conditions, diet etc.

Test conditions:

- details of test substance formulation including details of the physical form of the

material administered;
- details of the administration of the test substance including dosing volumes and time of dosing;
- details of food and water quality (including diet type/source, water source);
- the rationale for the selection of the starting dose.

Results:

- tabulation of response data and dose level for each animal (i.e. animals showing signs of toxicity including mortality; nature, severity, and duration of effects);
- tabulation of body weight and body weight changes;
- individual weights of animals at the day of dosing, in weekly intervals thereafter, and at the time of death or sacrifice;
- date and time of death if prior to scheduled sacrifice;
- time course of onset of signs of toxicity, and whether these were reversible for each animal;
- necropsy findings and histopathological findings for each animal, if available.

Discussion and interpretation of results.

Conclusions.

LITERATURE

(1) Roll R., Höfer-Bosse Th. And Kayser D. (1986). New Perspectives in Acute Toxicity Testing of Chemicals. Toxicol. Lett., Suppl. 31, 86.

(2) Roll R., Riebschläger M., Mischke U. and Kayser D. (1989). Neue Wege zur Bestimmung der akuten Toxizität von Chemikalien. Bundesgesundheitsblatt 32, 336-341.

(3) Diener W., Sichha L., Mischke U., Kayser D. and Schlede E. (1994). The Biometric Evaluation of the Acute-Toxic-Class Method (Oral). Arch. Toxicol. 68, 559-610.

(4) Diener W., Mischke U., Kayser D. and Schlede E. (1995). The Biometric Evaluation of the OECD Modified Version of the Acute-Toxic-Class Method (Oral). Arch. Toxicol. 69, 729-734.

(5) Diener W., and Schlede E. (1999) Acute Toxicity Class Methods: Alternatives to LD/LC50 Tests. ALTEX 16, 129-134.

(6) Schlede E., Mischke U., Roll R. and Kayser D. (1992). A National Validation Study of the Acute-Toxic-Class Method – An Alternative to the LD50 Test. Arch. Toxicol. 66, 455-470.

(7) Schlede E., Mischke U., Diener W. and Kayser D. (1994). The International Validation Study of the Acute-Toxic-Class Method (Oral). Arch. Toxicol. 69, 659-670.

(8) OECD (2000) Guidance Document on Acute Oral Toxicity. Environmental Health and Safety Monograph Series on Testing and Assessment No 24.

(9) OECD (2000) Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation Environmental Health and Safety Monograph Series on Testing and Assessment No 19.

(10) OECD (1998) Harmonized Integrated Hazard Classification System For Human Health And Environmental Effects Of Chemical Substances as endorsed by the 28th Joint Meeting of the

Chemicals Committee and the Working Party on Chemicals in November 1998, Part 2, p. 11 [http://webnet1.oecd.org/oecd/pages/home/displaygeneral/0,3380,EN-documents-521-14-no-24-no-0,FF.html].

(11) Lipnick R L, Cotruvo, J A, Hill R N, Bruce R D, Stitzel K A, Walker A P, Chu I; Goddard M, Segal L, Springer J A and Myers R C (1995) Comparison of the Up-and Down, Conventional LD_{50}, and Fixed Dose Acute Toxicity Procedures. Fd. Chem. Toxicol 33, 223-231.

(12) Chan P.K. and A.W. Hayes. (1994). Chap. 16. Acute Toxicity and Eye Irritancy. *Principles and Methods of Toxicology*. Third Edition. A.W. Hayes, Editor. Raven Press, Ltd., New York, USA.

423 OECD/OCDE

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.

Delayed death means that an animal does not die or appear moribund within 48 hours but dies later during the 14-day observation period.

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

GHS: Globally Harmonised Classification System for Chemical Substances and Mixtures. A joint activity of OECD (human health and the environment), UN Committee of Experts on Transport of Dangerous Goods (physical–chemical properties) and ILO (hazard communication) and co-ordinated by the Interorganisation Programme for the Sound Management of Chemicals (IOMC).

Impending death: when moribund state or death is expected prior to the next planned time of observation. Signs indicative of this state in rodents could include convulsions, lateral position, recumbence, and tremor (See the Humane Endpoint Guidance Document (9) for more details).

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 (2000 or 5000 mg/kg).

Moribund status: being in a state of dying or inability to survive, even if treated (See the Humane Endpoint Guidance Document (9) for more details).

Predictable death: presence of clinical signs indicative of death at a known time in the future before the planned end of the experimen ; for example: inability to reach water or food. (See the Humane Endpoint Guidance Document (9) for more details).

OECD/OCDE 423

ANNEX 2

PROCEDURE TO BE FOLLOWED FOR EACH OF THE STARTING DOSES

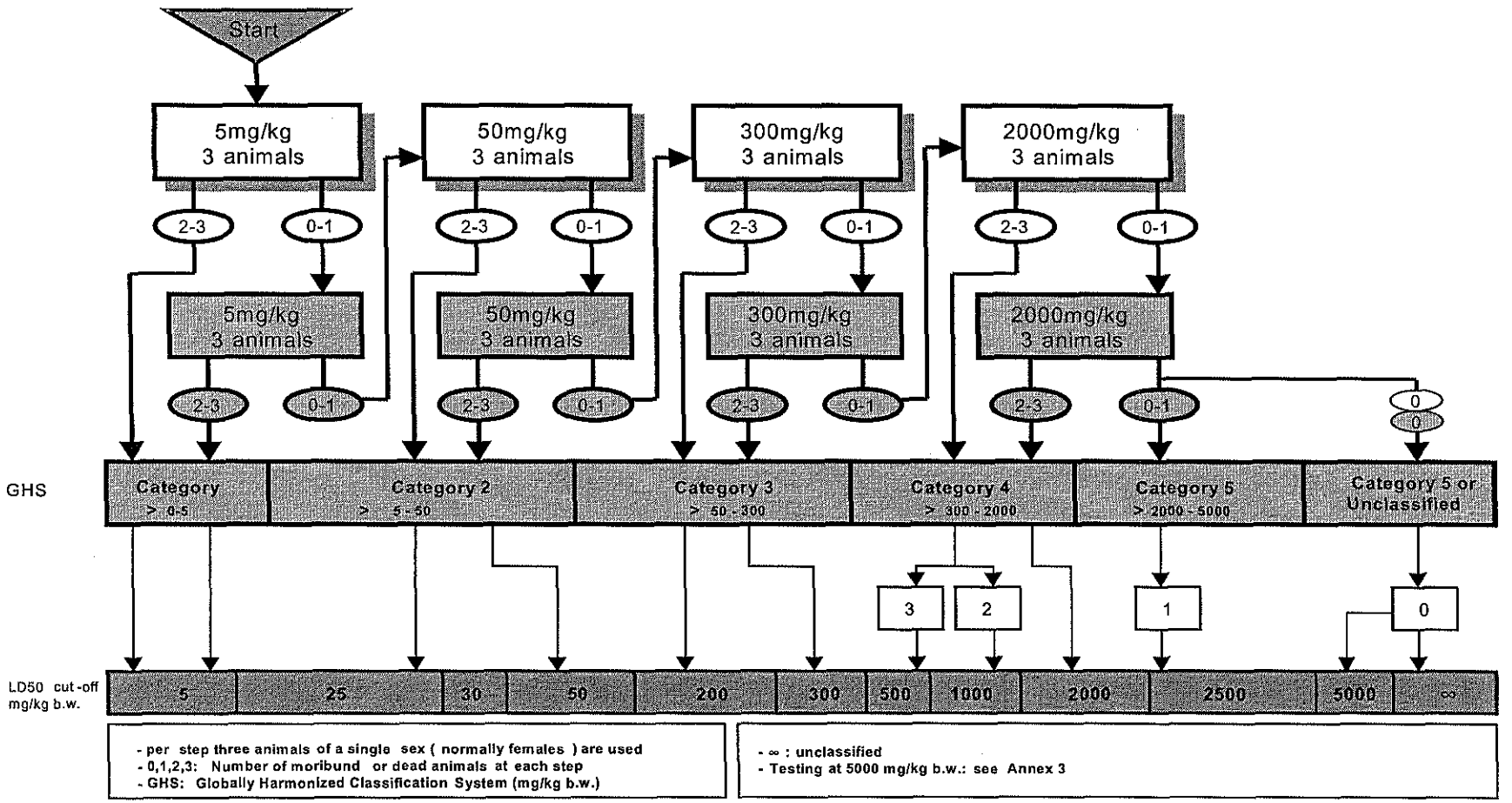
GENERAL REMARKS

1. For each starting dose, the respective testing schemes as included in this Annex outline the procedure to be followed.

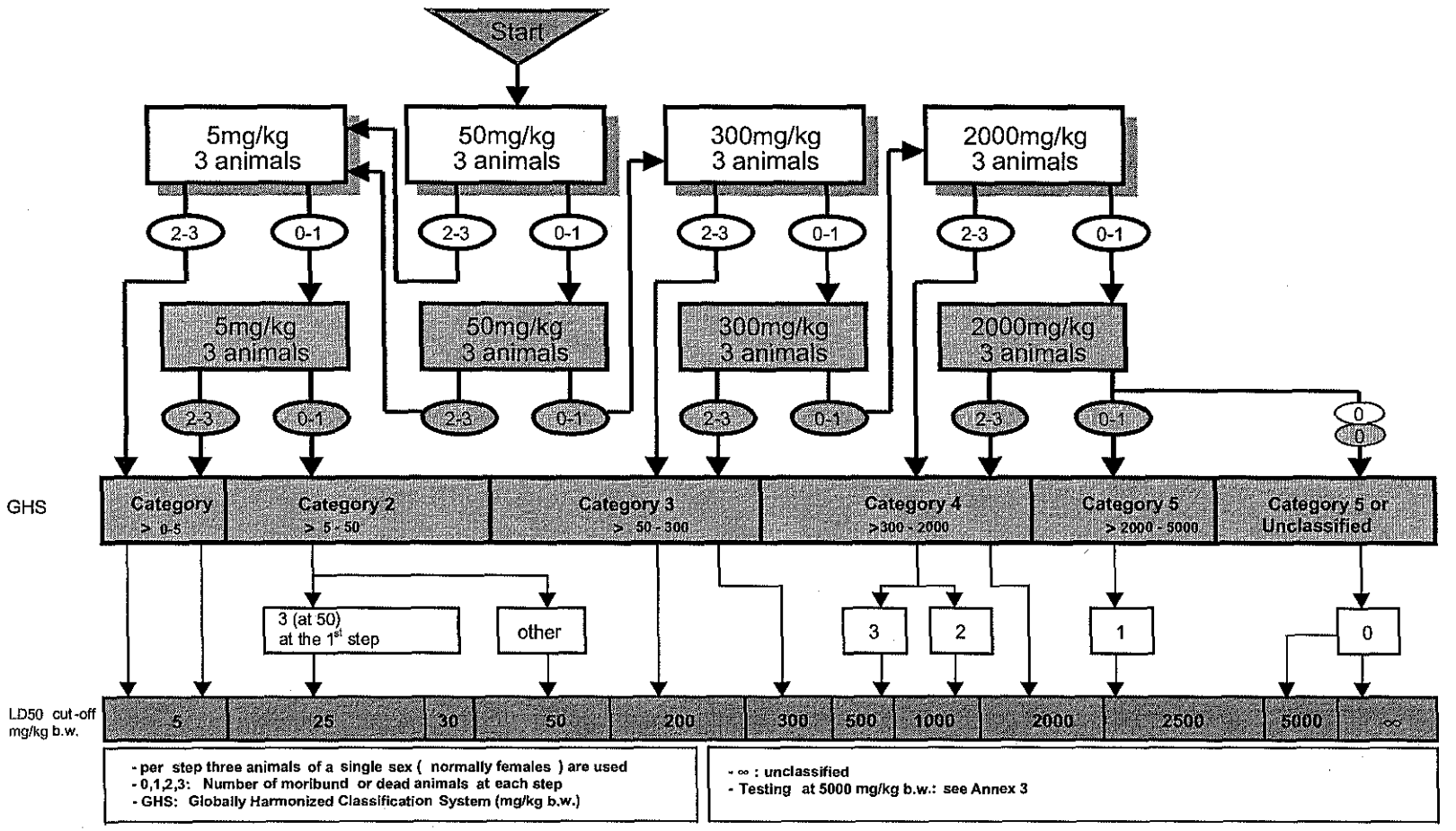
- Annex 2 a: Starting dose is 5 mg/kg bw
- Annex 2 b: Starting dose is 50 mg/kg bw
- Annex 2 c: Starting dose is: 300 mg/kg bw
- Annex 2 d: Starting dose is: 2000 mg/kg bw

Depending on the number of humanely killed or dead animals, the test procedure follows the indicated arrows.

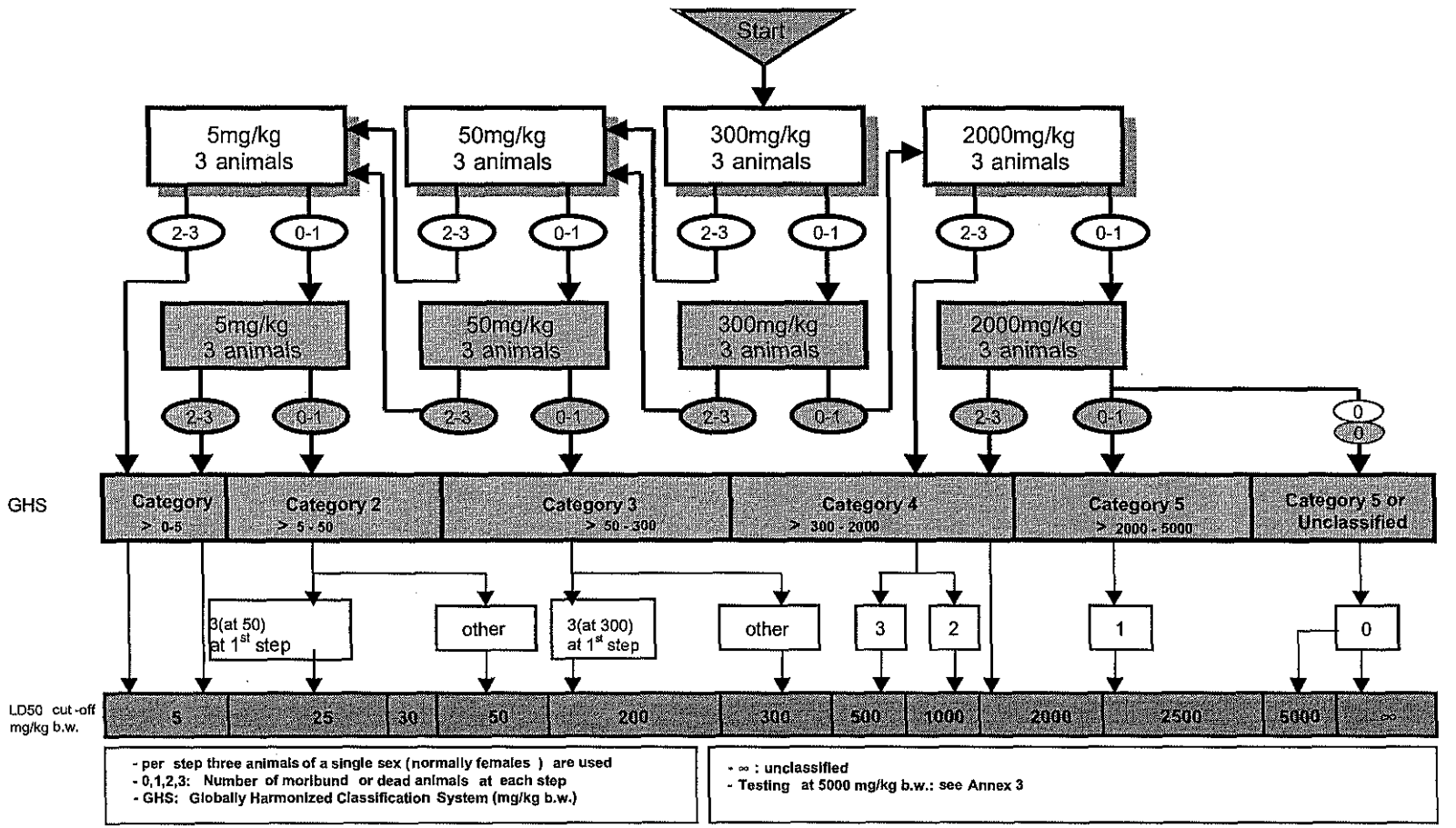
ANNEX 2a: TEST PROCEDURE WITH A STARTING DOSE OF 5 MG/KG BODY WEIGHT



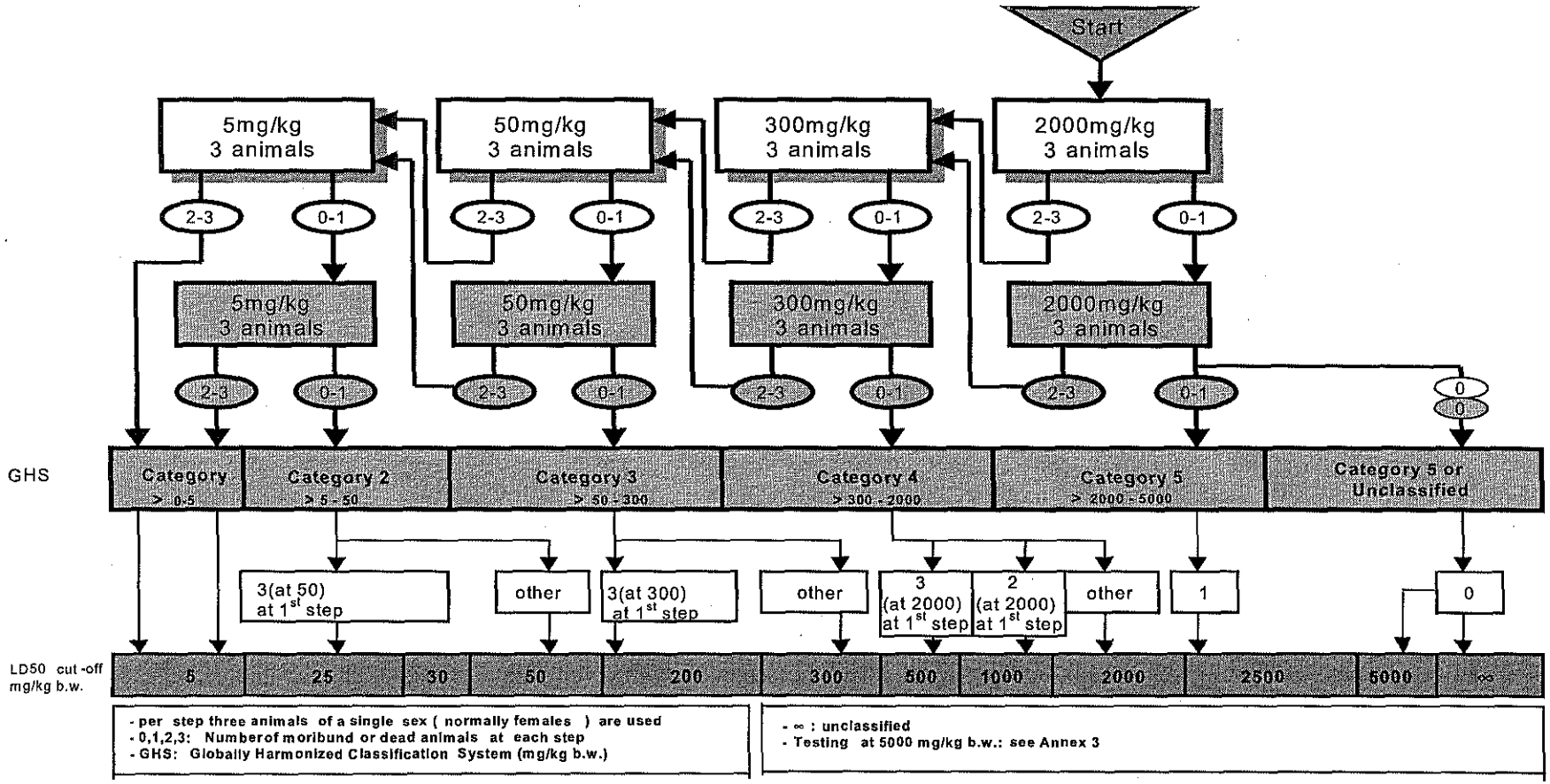
ANNEX 2b: TEST PROCEDURE WITH A STARTING DOSE OF 50 MG/KG BODY WEIGHT



ANNEX 2c: TEST PROCEDURE WITH A STARTING DOSE OF 300 MG/KG BODY WEIGHT



ANNEX 2d: TEST PROCEDURE WITH A STARTING DOSE OF 2000 MG/KG BODY WEIGHT



423 OECD/OCDE

ANNEX 3

CRITERIA FOR CLASSIFICATION OF TEST SUBSTANCES WITH EXPECTED LD50 VALUES EXCEEDING 2000 MG/KG WITHOUT THE NEED FOR TESTING

1. Criteria for hazard Category 5 are intended to enable the identification of test substances which are of relatively low acute toxicity hazard but which, under certain circumstances may present a danger to vulnerable populations. These substances are anticipated to have an oral or dermal LD50 in the range of 2000-5000 mg/kg or equivalent doses for other routes. The test substance should be classified in the hazard category defined by: 2000mg/kg<LD50<5000mg/kg (Category 5 in the GHS) in the following cases:

a) If directed to this category by any of the testing schemes of Annex 2a-2d, based on mortality incidences;

b) if reliable evidence is already available that indicates the LD50 to be in the range of Category 5 values, or other animal studies or toxic effects in humans indicate a concern for human health of an acute nature.

c) Through extrapolation, estimation or measurement of data if assignment to a more hazardous category is not warranted, and

- reliable information is available indicating significant toxic effects in humans, or
- any mortality is observed when tested up to Category 4 values by the oral route, or
- where expert judgement confirms significant clinical signs of toxicity, when tested up to Category 4 values, except for diarrhoea, piloerection or an ungroomed appearance, or
- where expert judgement confirms reliable information indicating the potential for significant acute effects from the other animal studies.

TESTING AT DOSES ABOVE 2000 MG/KG

2. Recognising the need to protect animal welfare, testing of animals in Category 5 (5000 mg/kg) ranges is discouraged and should only be considered when there is a strong likelihood that results of such a test have a direct relevance for protecting human or animal health (10). No further testing should be conducted at higher dose levels.

3. When testing is required a dose of 5000mg/kg, only one step (i.e. three animals) is required. If the first animal dosed dies , then dosing procedes at 2000mg/kg in accordance with the flow charts in Annex 2. If the first animal survives, two further animals are dosed. If only one of the three animal dies , the LD50 value is expected to exceed 5000mg/kg. If both animals die, then dosing proceeds at 2000mg/kg.

APPENDIX B4

OECD GUIDELINE 420: ACUTE ORAL TOXICITY – FIXED DOSE PROCEDURE

[This Page Intentionally Left Blank]

OECD/OCDE **420**

Adopted:
17th December 2001

OECD GUIDELINE FOR TESTING OF CHEMICALS

Acute Oral Toxicity – Fixed Dose Procedure

INTRODUCTION

1. OECD Guidelines for the Testing of Chemicals are periodically reviewed in the light of scientific progress or changing assessment practices. The original Guideline 420 was adopted in July 1992 as the first alternative to the conventional acute toxicity test, described in Test Guideline 401. Based on the recommendations of several expert meetings, revision was considered timely because: i) international agreement had been reached on harmonised LD50 cut-off values for the classification of chemical substances, which differ from the cut-offs recommended in the 1992 version of the Guideline, and ii) testing in one sex (usually females) is now considered sufficient.

2. Traditional methods for assessing acute toxicity use death of animals as an endpoint. In 1984, a new approach to acute toxicity testing was suggested by the British Toxicology Society based on the administration at a series of fixed dose levels (1). The approach avoided using death of animals as an endpoint, and relied instead on the observation of clear signs of toxicity at one of a series of fixed dose levels. Following UK (2) and international (3) *in vivo* validation studies the procedure was adopted by the Council as a Test Guideline in 1992. Subsequently, the statistical properties of the Fixed Dose Procedure have been evaluated using mathematical models in a series of studies (4)(5)(6). Together, the *in vivo* and modelling studies have demonstrated that the procedure is reproducible, uses fewer animals and causes less suffering than the traditional methods and is able to rank substances in a similar manner to the other acute toxicity testing methods (Test Guidelines 423 and 425).

3. Guidance on the selection of the most appropriate test method for a given purpose can be found in the Guidance Document on Acute Oral Toxicity Testing (7). This Guidance Document also contains additional information on the conduct and interpretation of Guideline 420.

4. Definitions used in the context of this Guideline are set out in Annex 1.

INITIAL CONSIDERATIONS

5. It is a principle of the method that in the main study only moderately toxic doses are used, and that administration of doses that are expected to be lethal should be avoided. Also, doses that are known to cause marked pain and distress, due to corrosive or severely irritant actions, need not be administered. Moribund animals, or animals obviously in pain or showing signs of severe and enduring distress shall be humanely killed, and are considered in the interpretation of the test results in the same way as animals that died on test. Criteria for making the decision to kill moribund or severely suffering animals, and guidance on the recognition of predictable or impending death, are the subject of a separate Guidance Document (8).

6. The method provides information on the hazardous properties and allows the substance to be ranked and classified according to the Globally Harmonised System (GHS) for the classification of chemicals which cause acute toxicity (9).

7. The testing laboratory should consider all available information on the test substance prior to conducting the study. Such information will include the identity and chemical structure of the substance; its physico-chemical properties; the results of any other *in vitro* or *in vivo* toxicity tests on the substance; toxicological data on structurally related substances; and the anticipated use(s) of the substance. This information is necessary to satisfy all concerned that the test is relevant for the protection of human health, and will help in the selection of an appropriate starting dose.

PRINCIPLE OF THE TEST

8. Groups of animals of a single sex are dosed in a stepwise procedure using the fixed doses of 5, 50, 300 and 2000 mg/kg (exceptionally an additional fixed dose of 5000 mg/kg may be considered, see paragraph 19). The initial dose level is selected on the basis of a sighting study as the dose expected to produce some signs of toxicity without causing severe toxic effects or mortality. Clinical signs and conditions associated with pain, suffering, and impending death, are described in detail in a separate OECD Guidance Document (8). Further groups of animals may be dosed at higher or lower fixed doses, depending on the presence or absence of signs of toxicity or mortality. This procedure continues until the dose causing evident toxicity or no more than one death is identified, or when no effects are seen at the highest dose or when deaths occur at the lowest dose.

DESCRIPTION OF THE METHOD

Selection of animal species

9. The preferred rodent species is the rat, although other rodent species may be used. Normally females are used (7). This is because literature surveys of conventional LD50 tests show that usually there is little difference in sensitivity between the sexes, but in those cases where differences are observed, females are generally slightly more sensitive (10). However, if knowledge of the toxicological or toxicokinetic properties of structurally related chemicals indicates that males are likely to be more sensitive then this sex should be used. When the test is conducted in males, adequate justification should be provided.

10. Healthy young adult animals of commonly used laboratory strains should be employed. Females should be nulliparous and non-pregnant. Each animal, at the commencement of its dosing, should be between 8 and 12 weeks old and its weight should fall in an interval within $\pm$ 20 % of the mean weight of any previously dosed animals.

Housing and feeding conditions

11. The temperature of the experimental animal room should be 22°C ($\pm$ 3°C). Although the relative humidity should be at least 30% and preferably not exceed 70% other than during room cleaning the aim should be 50-60%. Lighting should be artificial, the sequence being 12 hours light, 12 hours dark. For feeding, conventional laboratory diets may be used with an unlimited supply of drinking water. Animals may be group-caged by dose, but the number of animals per cage must not interfere with clear observations of each animal.

Preparation of animals

12. The animals are randomly selected, marked to permit individual identification, and kept in their cages for at least 5 days prior to the start of dosing to allow for acclimatisation to the laboratory conditions.

Preparation of doses

13. In general test substances should be administered in a constant volume over the range of doses to be tested by varying the concentration of the dosing preparation. Where a liquid end product or mixture is to be tested however, the use of the undiluted test substance, ie at a constant concentration, may be more relevant to the subsequent risk assessment of that substance, and is a requirement of some regulatory authorities. In either case, the maximum dose volume for administration must not be exceeded. The maximum volume of liquid that can be administered at one time depends on the size of the test animal. In rodents, the volume should not normally exceed 1mL/100g of body weight: however in the case of aqueous solutions 2 mL/100g body weight can be considered. With respect to the formulation of the dosing preparation, the use of an aqueous solution/suspension/emulsion is recommended wherever possible, followed in order of preference by a solution/suspension/emulsion in oil (e.g. corn oil) and then possibly solution in other vehicles. For vehicles other than water the toxicological characteristics of the vehicle should be known. Doses must be prepared shortly prior to administration unless the stability of the preparation over the period during which it will be used is known and shown to be acceptable.

PROCEDURE

Administration of doses

14. The test substance is administered in a single dose by gavage using a stomach tube or a suitable intubation canula. In the unusual circumstance that a single dose is not possible, the dose may be given in smaller fractions over a period not exceeding 24 hours.

15. Animals should be fasted prior to dosing (e.g. with the rat, food but not water should be withheld over-night; with the mouse, food but not water should be withheld for 3-4 hours). Following the period of fasting, the animals should be weighed and the test substance administered. After the substance has been administered, food may be withheld for a further 3-4 hours in rats or 1-2 hours in mice. Where a dose is administered in fractions over a period of time, it may be necessary to provide the animals with food and water depending on the length of the period.

Sighting study

16. The purpose of the sighting study is to allow selection of the appropriate starting dose for the main study. The test substance is administered to single animals in a sequential manner following the flow charts in Annex 2. The sighting study is completed when a decision on the starting dose for the main study can be made (or if a death is seen at the lowest fixed dose).

17. The starting dose for the sighting study is selected from the fixed dose levels of 5, 50, 300 and 2000 mg/kg as a dose expected to produce evident toxicity based, when possible, on evidence from *in vivo* and *in vitro* data from the same chemical and from structurally related chemicals. In the absence of such information, the starting dose will be 300 mg/kg.

18. A period of at least 24 hours will be allowed between the dosing of each animal. All animals should be observed for at least 14 days.

19. Exceptionally, and only when justified by specific regulatory needs, the use of an additional upper fixed dose level of 5000 mg/kg may be considered (see Annex 4). For reasons of animal welfare concern, testing of animals in GHS Category 5 ranges (2000-5000mg/kg) is discouraged and should only be considered when there is a strong likelihood that results of such a test have a direct relevance for protecting human or animal health or the environment.

20. In cases where an animal tested at the lowest fixed dose level (5mg/kg) in the sighting study dies, the normal procedure is to terminate the study and assign the substance to GHS Category 1 (as shown in Annex 2). However, if further confirmation of the classification is required, an optional supplementary procedure may be conducted, as follows. A second animal is dosed at 5mg/kg. If this second animal dies, then GHS Category 1 will be confirmed and the study will be immediately terminated. If the second animal survives, then a maximum of three additional animals will be dosed at 5mg/kg. Because there will be a high risk of mortality, these animals should be dosed in a sequential manner to protect animal welfare. The time interval between dosing each animal should be sufficient to establish that the previous animal is likely to survive. If a second death occurs, the dosing sequence will be immediately terminated and no further animals will be dosed. Because the occurence of a second death (irrespective of the number of animals tested at the time of termination) falls into outcome A (2 or more deaths), the classification rule of Annex 3 at the 5mg/kg fixed dose is followed (Category 1 if there are 2 or more deaths or Category 2 if there is no more than 1 death).

Main study

Numbers of animals and dose levels

21. The action to be taken following testing at the starting dose level is indicated by the flow charts in Annex 3. One of three actions will be required; either stop testing and assign the appropriate hazard classification class, test at a higher fixed dose or test at a lower fixed dose. However, to protect animals, a dose level that caused death in the sighting study will not be revisited in the main study (see Annex 3). Experience has shown that the most likely outcome at the starting dose level will be that the substance can be classified and no further testing will be necessary.

22. A total of five animals of one sex will normally be used for each dose level investigated. The five animals will be made up of one animal from the sighting study dosed at the selected dose level together with an additional four animals (except, unusually, if a dose level used on the main study was not included in the sighting study).

23. The time interval between dosing at each level is determined by the onset, duration, and severity of toxic signs. Treatment of animals at the next dose should be delayed until one is confident of survival of the previously dosed animals. A period of 3 or 4 days between dosing at each dose level is recommended, if needed, to allow for the observation of delayed toxicity. The time interval may be adjusted as appropriate, e.g., in case of inconclusive response.

24. When the use of an upper fixed dose of 5000 mg/kg is considered, the procedure outlined in Annex 4 should be followed (see also paragraph 19).

Limit test

25. The limit test is primarily used in situations where the experimenter has information indicating that the test material is likely to be nontoxic, i.e., having toxicity only above regulatory limit doses. Information about the toxicity of the test material can be gained from knowledge about similar tested

compounds or similar tested mixtures or products, taking into consideration the identity and percentage of components known to be of toxicological significance. In those situations where there is little or no information about its toxicity, or in which the the test material is expected to be toxic, the main test should be performed.

26. Using the normal procedure, a sighting study starting dose of 2000mg/kg (or exceptionally 5000mg/kg) followed by dosing of a further four animals at this level serves as a limit test for this guideline.

OBSERVATIONS

27. Animals are observed individually after dosing at least once during the first 30 minutes, periodically during the first 24 hours, with special attention given during the first 4 hours, and daily thereafter, for a total of 14 days, except where they need to be removed from the study and humanely killed for animal welfare reasons or are found dead. However, the duration of observation should not be fixed rigidly. It should be determined by the toxic reactions, time of onset and length of recovery period, and may thus be extended when considered necessary. The times at which signs of toxicity appear and disappear are important, especially if there is a tendency for toxic signs to be delayed (11). All observations are systematically recorded, with individual records being maintained for each animal.

28. Additional observations will be necessary if the animals continue to display signs of toxicity. Observations should include changes in skin and fur, eyes and mucous membranes, and also respiratory, circulatory, autonomic and central nervous systems, and somatomotor activity and behaviour pattern. Attention should be directed to observations of tremors, convulsions, salivation, diarrhoea, lethargy, sleep and coma. The principles and criteria summarised in the Humane Endpoints Guidance Document should be taken into consideration (8). Animals found in a moribund condition and animals showing severe pain or enduring signs of severe distress should be humanely killed. When animals are killed for humane reasons or found dead, the time of death should be recorded as precisely as possible.

Body weight

29. Individual weights of animals should be determined shortly before the test substance is administered and at least weekly thereafter. Weight changes should be calculated and recorded. At the end of the test surviving animals are weighed and then humanely killed.

Pathology

30. All test animals (including those that die during the test or are removed from the study for animal welfare reasons) should be subjected to gross necropsy. All gross pathological changes should be recorded for each animal. Microscopic examination of organs showing evidence of gross pathology in animals surviving 24 or more hours after the initial dosing may also be considered because it may yield useful information.

DATA AND REPORTING

Data

31. Individual animal data should be provided. Additionally, all data should be summarised in tabular form, showing for each test group the number of animals used, the number of animals displaying signs of toxicity, the number of animals found dead during the test or killed for humane reasons, time of death of individual animals, a description and the time course of toxic effects and reversibility, and necropsy findings.

Test report

32. The test report must include the following information, as appropriate:

Test substance:

- physical nature, purity, and, where relevant, physico-chemical properties (including isomerisation);
- identification data, including CAS number.

Vehicle (if appropriate):

- justification for choice of vehicle, if other than water.

Test animals:

- species/strain used;
- microbiological status of the animals, when known;
- number, age and sex of animals (including, where appropriate, a rationale for use of males instead of females);
- source, housing conditions, diet etc.

Test conditions:

- details of test substance formulation, including details of the physical form of the material administered;
- details of the administration of the test substance including dosing volumes and time of dosing;
- details of food and water quality (including diet type/source, water source);
- the rationale for the selection of the starting dose.

Results:

- tabulation of response data and dose level for each animal (i.e. animals showing signs of toxicity including mortality, nature, severity and duration of effects);
- tabulation of body weight and body weight changes;
- individual weights of animals at the day of dosing, in weekly intervals thereafter, and at time of death or sacrifice;
- date and time of death if prior to scheduled sacrifice;
- time course of onset of signs of toxicity and whether these were reversible for each animal;
- necropsy findings and histopathological findings for each animal, if available.

Discussion and interpretation of results.

Conclusions.

OECD/OCDE 420

LITERATURE

(1) British Toxicology Society Working Party on Toxicity (1984). Special report: a new approach to the classification of substances and preparations on the basis of their acute toxicity. Human Toxicol., 3, 85-92.

(2) Van den Heuvel, M.J., Dayan, A.D. and Shillaker, R.O. (1987). Evaluation of the BTS approach to the testing of substances and preparations for their acute toxicity. Human Toxicol., 6, 279-291.

(3) Van den Heuvel, M.J., Clark, D.G., Fielder, R.J., Koundakjian, P.P., Oliver, G.J.A., Pelling, D., Tomlinson, N.J. and Walker, A.P. (1990). The international validation of a fixed-dose procedure as an alternative to the classical LD_{50} test. Fd. Chem. Toxicol. 28, 469-482.

(4) Whitehead, A. and Curnow, R.N. (1992). Statistical evaluation of the fixed-dose procedure. Fd. Chem. Toxicol., 30, 313-324.

(5) Stallard, N. and Whitehead, A. (1995). Reducing numbers in the fixed-dose procedure. Human Exptl. Toxicol. 14, 315-323.

(6) Stallard, N., Whitehead, A. and Ridgeway, P. (2000). Statistical evaluation of modifications to the fixed dose procedure (manuscript in preparation).

(7) OECD (2000). Guidance Document on Acute Oral Toxicity. Environmental Health and Safety Monograph Series on Testing and Assessment No.24.

(8) OECD (2000). Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation. Environmental Health and Safety Monograph Series on Testing and Assesment No 19.

(9) OECD (1998). Harmonised Integrated Hazard Classification for Human Health and Environmental Effects of Chemical Substances as endorsed by the 28th Joint Meeting of the Chemicals Committee and the Working Party on Chemicals in November 1998, Part 2, p.11 [http://webnet1.oecd.org/oecd/pages/home/displaygeneral/0,3380,EN-documents-521-14-no-24-no-0,FF.html].

(10) Lipnick, R.L., Cotruvo, J.A., Hill, R.N., Bruce, R.D., Stitzel, K.A., Walker, A.P., Chu, I., Goddard, M., Segal, L., Springer, J.A. and Myers, R.C. (1995). Comparison of the Up-and-Down, Conventional LD_{50}, and Fixed-Dose Acute Toxicity Procedures. Fd. Chem. Toxicol. 33, 223-231.

(11) Chan P.K and A.W. Hayes (1994) Chapter 16 Acute Toxicity and Eye Irritation In: Principles and Methods of Toxicology. 3rd Edition. A.W. Hayes , Editor. Raven Press, Ltd. New York, USA.

420 OECD/OCDE

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.

Delayed death means that an animal does not die or appear moribund within 48 hours but dies later during the 14-day observation period.

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

Evident toxicity is a general term describing clear signs of toxicity following the administration of test substance, (see Van den Heuvel, M.J., Clark, D.G., Fielder, R.J., Koundakjian, P.P., Oliver, G.J.A., Pelling, D., Tomlinson, N.J. and Walker, A.P. (1990). The international validation of a fixed-dose procedure as an alternative to the classical LD_{50} test. Fd. Chem. Toxicol. 28, 469-482. (3) for examples) such that at the next highest fixed dose either severe pain and enduring signs of severe distress, moribund status (criteria are presented in the Humane Endpoints Guidance Document (8), or probable mortality in most animals can be expected.

GHS: Globally Harmonised Classification System for Chemical Substances and Mixtures. A joint activity of OECD (human health and the environment), UN Committee of Experts on Transport of Dangerous Goods (physical–chemical properties) and ILO (hazard communication) and co-ordinated by the Interorganisation Programme for the Sound Management of Chemicals (IOMC).

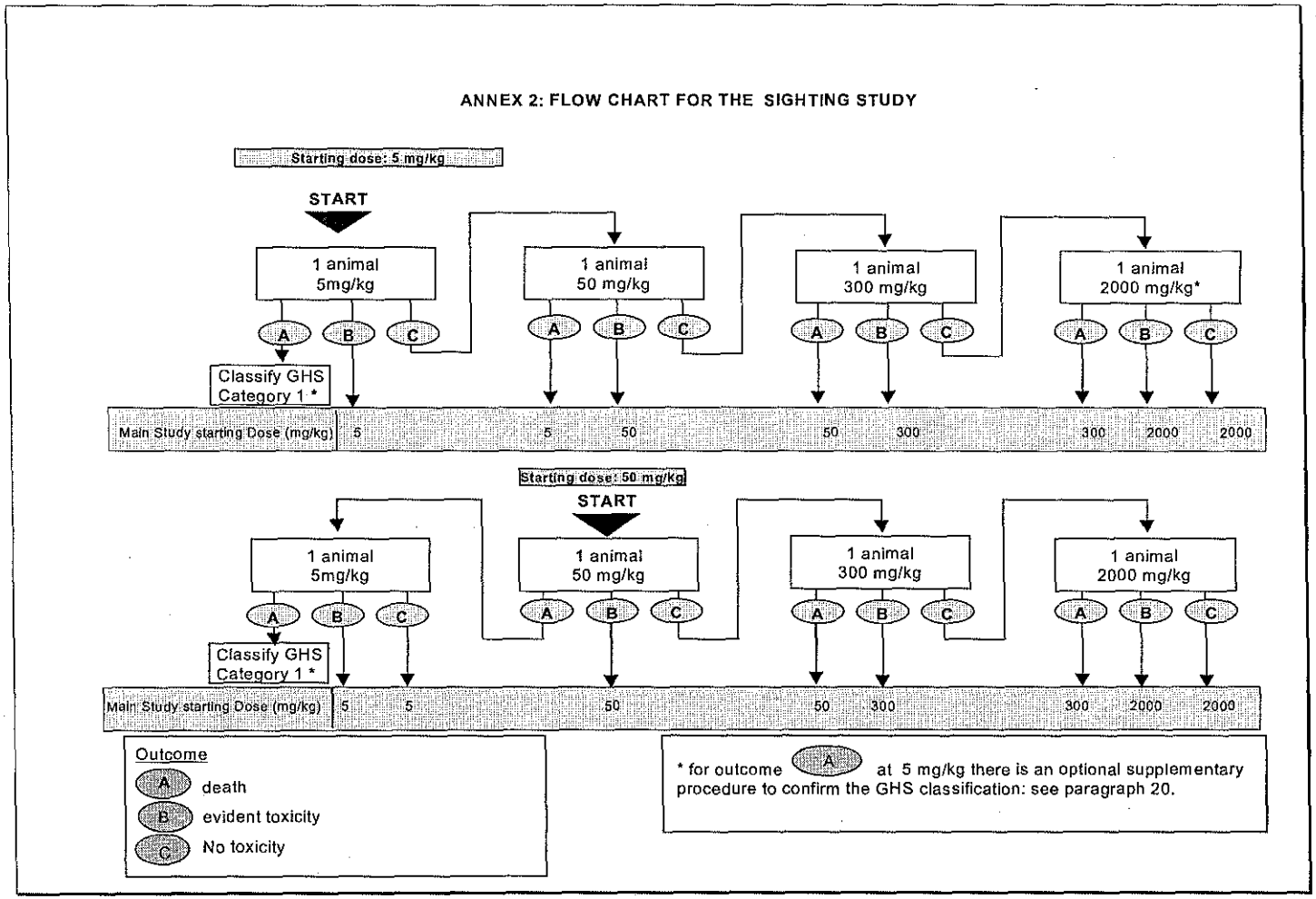
Impending death: when moribund state or death is expected prior to the next planned time of observation. Signs indicative of this state in rodents could include convulsions, lateral position, recumbence, and tremor. (See the Humane Endpoint Guidance Document (8) for more details).

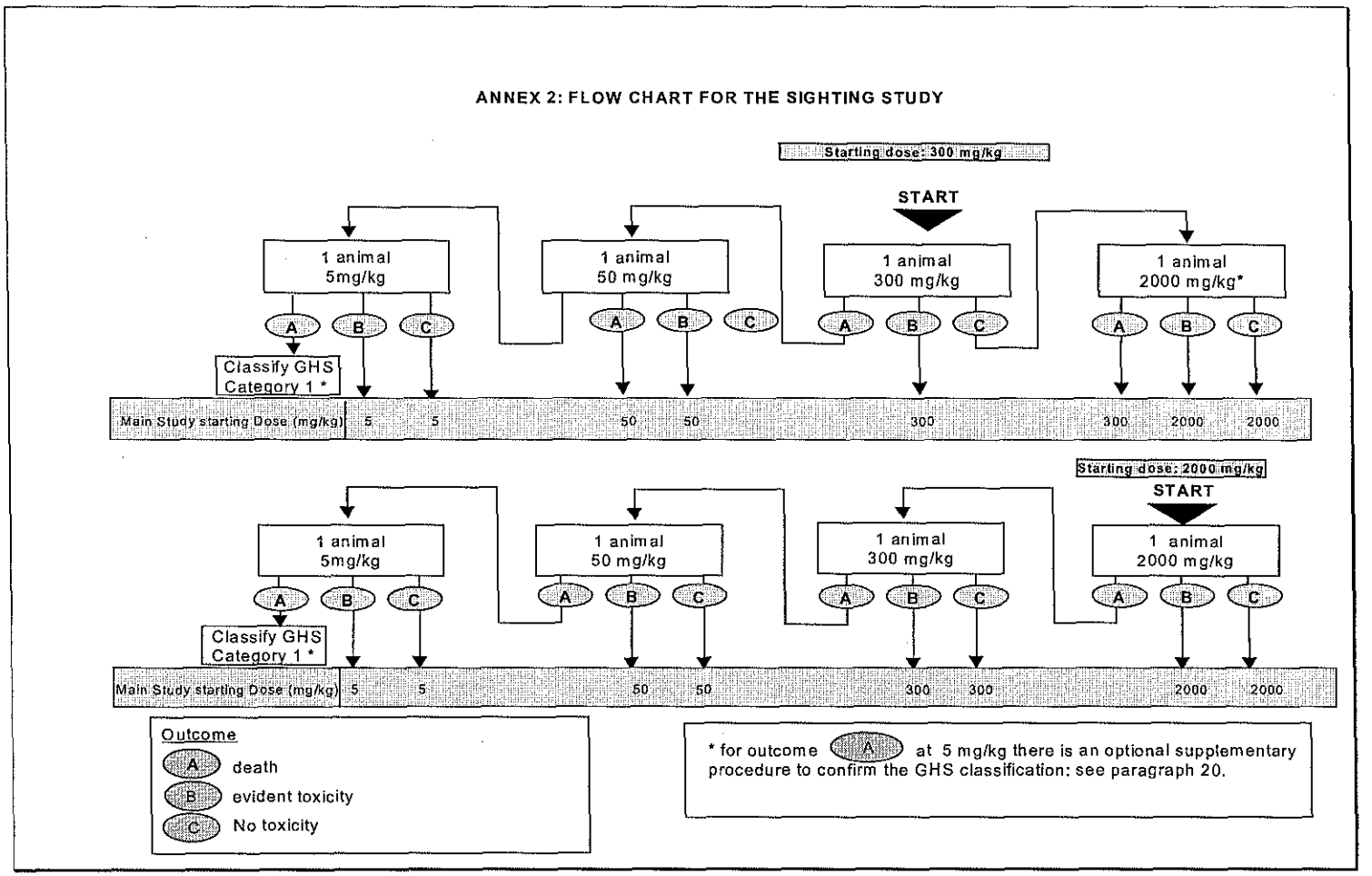
LD50 (median lethal oral dose) is a statistically derived single dose of a substance that can be expected to cause death in 50 per cent of animals when administered by the oral route. The LD50 value is expressed in terms of weight of test substance per unit weight of test animal (mg/kg).

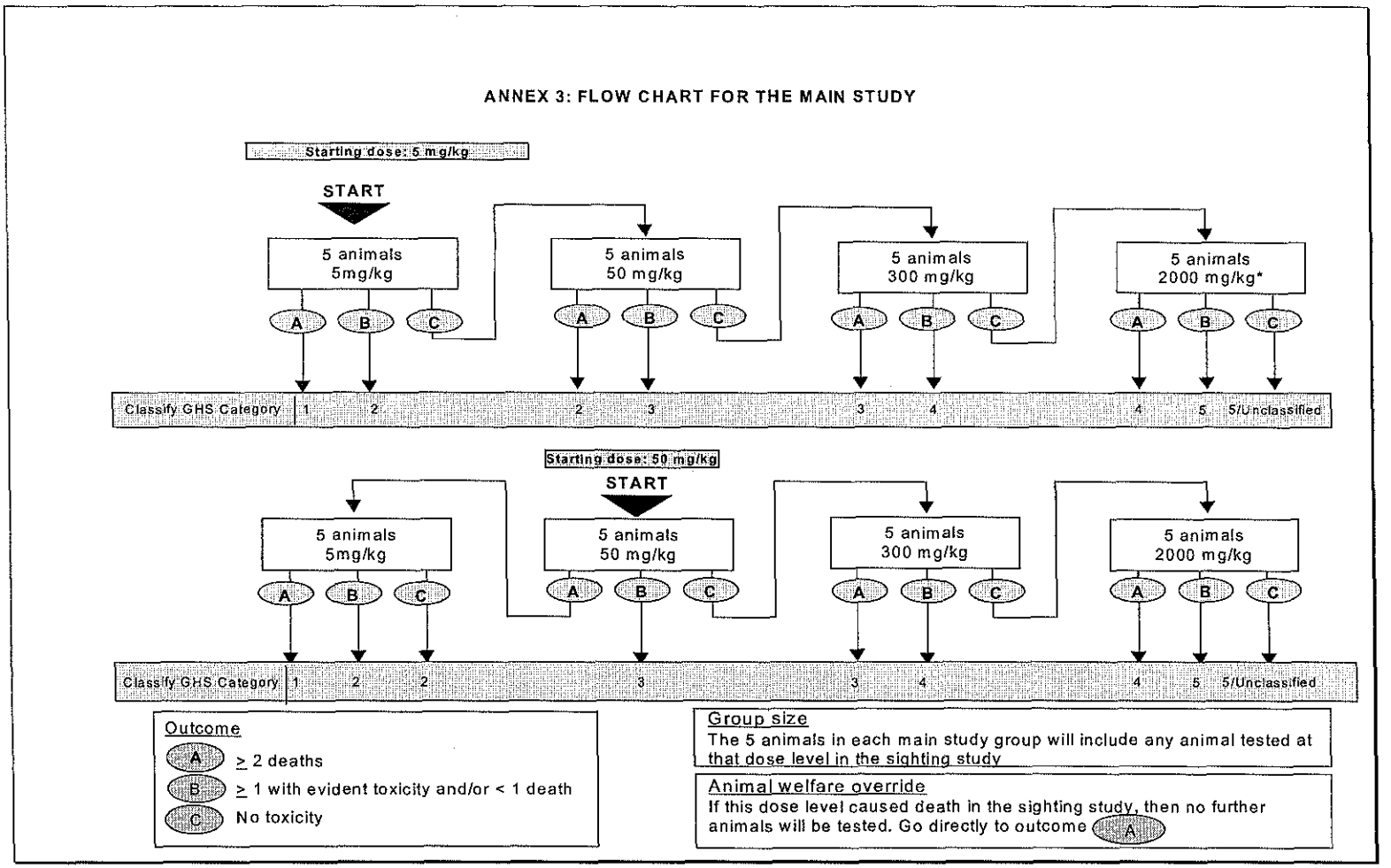
Limit dose refers to a dose at an upper limitation on testing (2000 or 5000 mg/kg).

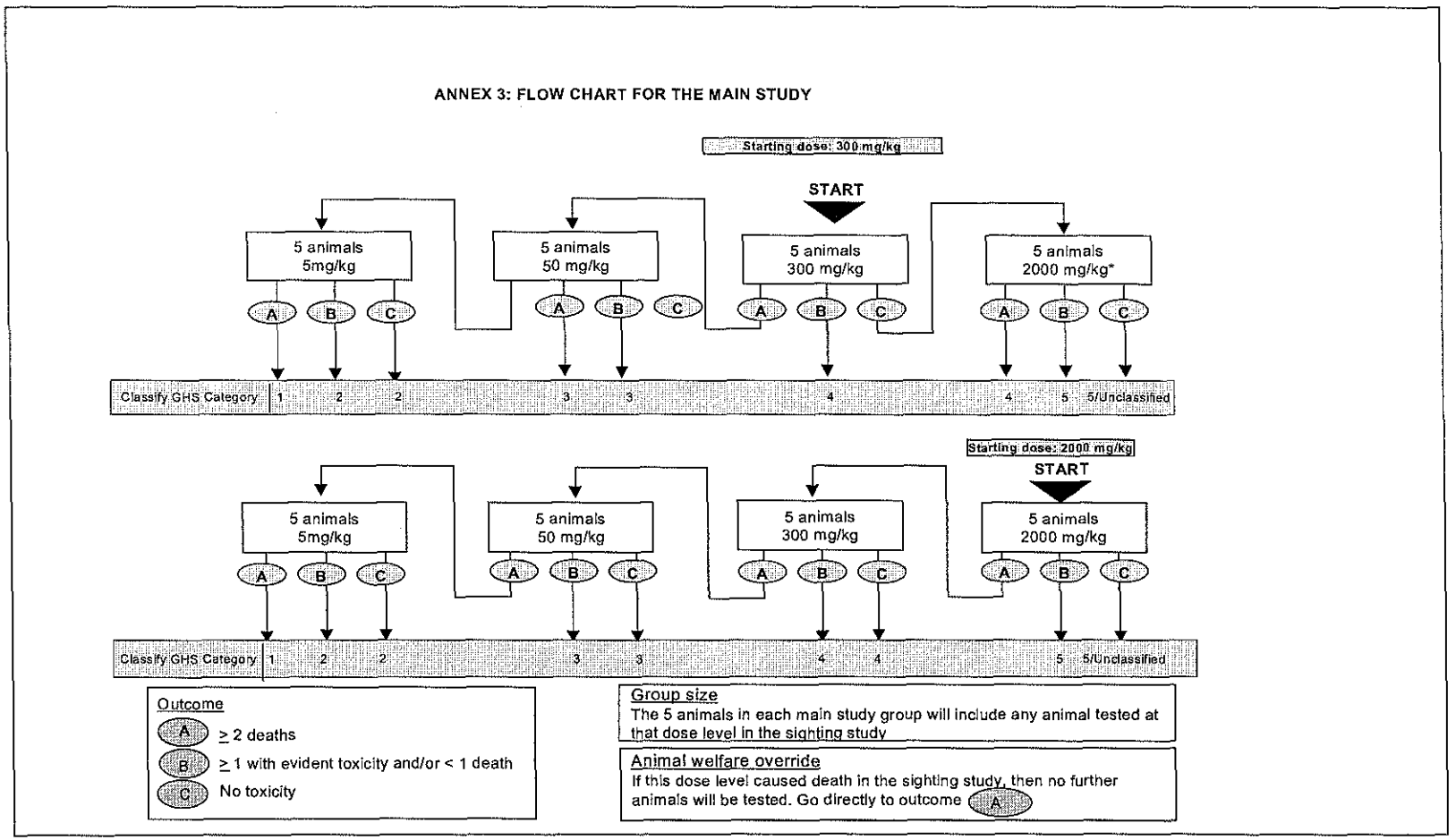
Moribund status: being in a state of dying or inability to survive, even if treated. (See the Humane Endpoint Guidance Document (8) for more details).

Predictable death: presence of clinical signs indicative of death at a known time in the future before the planned end of the experiment, for example: inability to reach water or food. (See the Humane Endpoint Guidance Document (8) for more details).


ANNEX 2: FLOW CHART FOR THE SIGHTING STUDY
Starting dose: 5 mg/kg
START
1 animal 5mg/kg
1 animal 50 mg/kg
1 animal 300 mg/kg
1 animal 2000 mg/kg*
A
B
C
Classify GHS Category 1 *
Main Study starting Dose (mg/kg)
5
5
50
50
300
300
2000
2000
Starting dose: 50 mg/kg
START
1 animal 5mg/kg
1 animal 50 mg/kg
1 animal 300 mg/kg
1 animal 2000 mg/kg
Classify GHS Category 1 *
Main Study starting Dose (mg/kg)
5
5
50
50
300
300
2000
2000
Outcome
A death
B evident toxicity
C No toxicity
* for outcome A at 5 mg/kg there is an optional supplementary procedure to confirm the GHS classification: see paragraph 20.


ANNEX 2: FLOW CHART FOR THE SIGHTING STUDY
Starting dose: 300 mg/kg
START
1 animal 5mg/kg
1 animal 50 mg/kg
1 animal 300 mg/kg
1 animal 2000 mg/kg*
A
B
C
Classify GHS Category 1 *
Main Study starting Dose (mg/kg) 5 5 50 50 300 300 2000 2000
Starting dose: 2000 mg/kg
START
1 animal 5mg/kg
1 animal 50 mg/kg
1 animal 300 mg/kg
1 animal 2000 mg/kg
Classify GHS Category 1 *
Main Study starting Dose (mg/kg) 5 5 50 50 300 300 2000 2000
Outcome
A death
B evident toxicity
C No toxicity
* for outcome A at 5 mg/kg there is an optional supplementary procedure to confirm the GHS classification: see paragraph 20.


ANNEX 3: FLOW CHART FOR THE MAIN STUDY
Starting dose: 5 mg/kg
START
5 animals 5mg/kg
5 animals 50 mg/kg
5 animals 300 mg/kg
5 animals 2000 mg/kg*
A
B
C
Classify GHS Category
1
2
2
3
3
4
4
5
5/Unclassified
Starting dose: 50 mg/kg
START
5 animals 5mg/kg
5 animals 50 mg/kg
5 animals 300 mg/kg
5 animals 2000 mg/kg
Classify GHS Category
1
2
2
3
3
4
4
5
5/Unclassified
Outcome
A ≥ 2 deaths
B ≥ 1 with evident toxicity and/or < 1 death
C No toxicity
Group size
The 5 animals in each main study group will include any animal tested at that dose level in the sighting study
Animal welfare override
If this dose level caused death in the sighting study, then no further animals will be tested. Go directly to outcome A


ANNEX 3: FLOW CHART FOR THE MAIN STUDY
Starting dose: 300 mg/kg
START
5 animals 5mg/kg
5 animals 50 mg/kg
5 animals 300 mg/kg
5 animals 2000 mg/kg*
A B C
Classify GHS Category 1 2 2 3 3 4 4 5 5/Unclassified
Starting dose: 2000 mg/kg
START
5 animals 5mg/kg
5 animals 50 mg/kg
5 animals 300 mg/kg
5 animals 2000 mg/kg
A B C
Classify GHS Category 1 2 2 3 3 4 4 5 5/Unclassified
Outcome
A ≥ 2 deaths
B ≥ 1 with evident toxicity and/or < 1 death
C No toxicity
Group size
The 5 animals in each main study group will include any animal tested at that dose level in the sighting study
Animal welfare override
If this dose level caused death in the sighting study, then no further animals will be tested. Go directly to outcome A

OECD/OCDE 420

ANNEX 4

CRITERIA FOR CLASSIFICATION OF TEST SUBSTANCES WITH EXPECTED LD50 VALUES EXCEEDING 2000 MG/KG WITHOUT THE NEED FOR TESTING.

1. Criteria for hazard Category 5 are intended to enable the identification of test substances which are of relatively low acute toxicity hazard but which, under certain circumstances may present a danger to vulnerable populations. These substances are anticipated to have an oral or dermal LD50 in the range of 2000-5000 mg/kg or equivalent doses for other routes. Test substances could be classified in the hazard category defined by: 2000mg/kg <LD50 < 5000mg/kg (Category 5 in the GHS) in the following cases:

a) if directed to this category by any of the testing schemes of Annex 3, based on mortality incidences;

b) if reliable evidence is already available that indicates the LD50 to be in the range of Category 5 values; or other animal studies or toxic effects in humans indicate a concern for human health of an acute nature;

c) through extrapolation, estimation or measurement of data if assignment to a more hazardous category is not warranted and
   - reliable information is available indicating significant toxic effects in humans, or
   - any mortality is observed when tested up to category 4 values by the oral route, or
   - where expert judgement confirms significant clinical signs of toxicity, when tested up to Category 4 values, except for diarrhoea, piloerection or an ungroomed appearance, or
   - where expert judgement confirms reliable information indicating the potential for significant acute effects from the other animal studies.

TESTING AT DOSES ABOVE 2000 MG/KG

2. Exceptionally, and only when justified by specific regulatory needs, the use of an additional upper fixed dose level of 5000 mg/kg may be considered. Recognising the need to protect animal welfare, testing at 5000 mg/kg is discouraged and should only be considered when there is a strong likelihood that the results of such a test would have a direct relevance for protecting animal or human health (9).

Sighting Study

3. The decision rules governing the sequential procedure presented in Annex 2 are extended to include a 5000 mg/kg dose level. Thus, when a sighting study starting dose of 5000 mg/kg is used outcome A (death) will require a second animal to be tested at 2000 mg/kg; outcomes B and C (evident toxicity or no toxicity) will allow the selection of 5000 mg/kg as the main study starting dose. Similarly, if a starting dose other than 5000 mg/kg is used then testing will progress to 5000 mg/kg in the event of outcomes B or C at 2000 mg/kg; a subsequent 5000 mg/kg outcome A will dictate a main study starting dose of 2000 mg/kg and outcomes B and C will dictate a main study starting dose of 5000 mg/kg.

Main Study

4. The decision rules governing the sequential procedure presented in Annex 3 are extended to include a 5000 mg/kg dose level. Thus, when a main study starting dose of 5000 mg/kg is used, outcome A (≥2 deaths) will require the testing of a second group at 2000 mg/kg; outcome B (evident toxicity and/or ≤1 death) or C (no toxicity) will result in the substance being unclassified according to GHS.

420 **OECD/OCDE**

Similarly, if a starting dose other than 5000 mg/kg is used then testing will progress to 5000 mg/kg in the event of outcome C at 2000 mg/kg; a subsequent 5000 mg/kg outcome A will result in the substance being assigned to GHS Category 5 and outcomes B or C will lead to the substance being unclassified.

APPENDIX B5

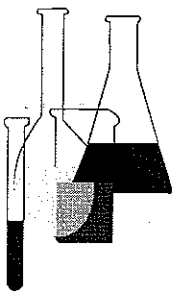
Health Effects Test Guidelines OPPTS 870.1100: Acute Oral Toxicity

[This Page Intentionally Left Blank]

United States Environmental Protection Agency | Prevention, Pesticides and Toxic Substances (7101) | EPA 712–C–02–190 December 2002

EPA

Health Effects Test Guidelines OPPTS 870.1100 Acute Oral Toxicity



INTRODUCTION

This guideline is one of a series of test guidelines that have been developed by the Office of Prevention, Pesticides and Toxic Substances, United States Environmental Protection Agency for use in the testing of pesticides and toxic substances, and the development of test data that must be submitted to the Agency for review under Federal regulations.

The Office of Prevention, Pesticides and Toxic Substances (OPPTS) has developed this guideline through a process of harmonization that blended the testing guidance and requirements that existed in the Office of Pollution Prevention and Toxics (OPPT) and appeared in Title 40, Chapter I, Subchapter R of the Code of Federal Regulations (CFR), the Office of Pesticide Programs (OPP) which appeared in publications of the National Technical Information Service (NTIS) and the guidelines published by the Organization for Economic Cooperation and Development (OECD).

The purpose of harmonizing these guidelines into a single set of OPPTS guidelines is to minimize variations among the testing procedures that must be performed to meet the data requirements of the U. S. Environmental Protection Agency under the Toxic Substances Control Act (15 U.S.C. 2601) and the Federal Insecticide, Fungicide and Rodenticide Act (7 U.S.C. 136, *et seq.*).

Final Guideline Release: This guideline is available from the U.S. Government Printing Office, Washington, DC 20402 on disks or paper copies: call (202) 512–0132. This guideline is also available electronically in PDF (portable document format) from EPA's Internet Web site at http://www.epa.gov/opptsfrs/home/guidelin.htm. Also, the Agency has developed, and strongly recommends users to solely use, the software program for performing the Up-and-Down Procedure and calculating the LD50 and confidence interval. The software program (AOT425StatPgm) is available on EPA's Internet Web site at http://www.epa.gov/oppfead1/harmonized.

OPPTS 870.1100 Acute oral toxicity.

(a) **Scope—Applicability**. This guideline is intended to meet testing requirements of both the Federal Insecticide, Fungicide, and Rodenticida Act (FIFRA) (7 U.S.C. 136, *et seq.*) and the Toxic Substances Control Act (TSCA) (15 U.S.C. 2601).

(2) **Background**. The source material for this revised harmonized test guideline is OPPTS 870.1100 Acute Oral Toxicity, dated August 1998 and OECD test Guideline 425 Acute Oral Toxicity–Up-and-Down Procedure.

(b) **Purpose**. In the assessment and evaluation of the toxic characteristics of a substance, determination of acute oral toxicity is usually an initial step. It provides information on health and environmental hazards likely to arise from short-term exposure by the oral route. Data from an acute study may serve as a basis for classification and labeling. It is traditionally a step in establishing a dosage regimen in subchronic and other studies and may provide initial information on the mode of toxic action of a substance. An evaluation of acute toxicity data should include the relationship, if any, between the exposure of animals to the test substance and the incidence and severity of all abnormalities, including behavioral and clinical abnormalities, the reversibility of observed abnormalities, gross lesions, body weight changes, effects on mortality, and any other toxic effects.

(c) **Definitions**. The definitions in Section 3 of the Toxic Substances Control Act (TSCA) and the definitions in 40 CFR Part 792—Good Laboratory Practice Standards apply to this test guideline. The following definitions also apply to this test guideline.

Acute oral toxicity is the adverse effects occurring within a short time of oral administration of a single dose of a substance or multiple doses given within 24 hours.

Confidence interval (CI) is an interval estimate, a range of values, intended to include the true LD_{50} with a specified degree of confidence.

Delayed death means that an animal does not die or appear moribund within 48 hours, but dies later during the 14-day observation period.

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

Dose progression factor, sometimes termed a dose spacing factor, refers to the multiple by which a dose is increased (i.e., the dose progression) when an animal survives or the divisor by which it is decreased when an animal dies. The dose progression factor is recommended to be the antilog of 1/(the estimated slope of the dose-response curve). The default

dose progression factor is recommended to be 3.2 = antilog 0.5 = antilog (1/2).

LD_{50} (median lethal dose), oral, 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 LD_{50} 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 (2000–5000 mg/kg).

Moribund status of an animal refers to being in a state of dying or inability to survive, even if treated.

Nominal sample size refers to the total number of tested animals, reduced by one less than the number of like responses at the beginning of the series, or by the number of tested animals up to but not including the pair that creates the first reversal. For example, for a series where X and O indicate opposite animal outcomes (for instance, X could be dies within 48 hours and O survives) in a pattern as follows: OOOXXOXO, we have the total number of tested animals (or sample size in the conventional sense) as 8 and the nominal sample size as 6. This particular example shows 4 animals following a reversal. It is important to note whether a count in a particular part of the guideline refers to the nominal sample size or to the total number tested. For example, the maximum actual number tested is 15. When testing is stopped based on that basis, the nominal sample size will be less than or equal to 15. Members of the nominal sample start with the (r-1)st animal (the animal before the second in the reversal pair) (see reversal below).

Probit is an abbreviation for the term "*prob*ability *i*ntegral *t*ransformation" 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 LD_{50} or median response.

Reversal is a situation where nonresponse is observed at some dose, and a response is observed at the next dose tested, or vice versa (i.e., response followed by nonresponse). Thus, a reversal is created by a pair of responses. The first such pair occurs at animals numbered r-1 and r.

Sigma is the standard deviation of a log normal curve describing the range of tolerances of test subjects to the chemical (where a subject is expected capable of responding if the chemical dose exceeds the subject's tolerance). The estimated *sigma* provides an estimate of the variation

among test animals in response to a full range of doses. See slope and probit.

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*.

Stopping rule is used in this guideline synonymously with (1) a specific stopping criterion and (2) the collection of all criteria determining when a testing sequence terminates. In particular, for the main test, stopping rule is used in paragraph (e)(2)(ii) of this guideline as a shorthand for the criterion that relies on comparison of ratios to a critical value.

(d) **Approaches to the determination of acute toxicity**. EPA recommends the Up-and-Down Procedure (UDP) as detailed in this guideline and adopted by the Organization for Economic Cooperation and Development (OECD) as test Guideline 425 (see paragraph (n)(1) of this guideline), to assess acute oral toxicity. This method provides a point estimate of lethality and confidence interval around the LD50. Acute oral toxicity testing may also be performed using the Fixed Dose Method of OECD Guideline 420 (see paragraph (n)(2) of this guideline) or the Acute Toxic Class Method of OECD Guideline 423 (see paragraph (n)(3) of this guideline). These methods assess lethality within a dose range.

(e) **Introduction to the UDP**—(1) **Background**. (i) The concept of the up-and-down testing approach was first described by Dixon and Mood (see paragraphs (n)(4) through (n)(7) of this guideline). In 1985, Bruce proposed to use an UDP for the determination of acute toxicity of chemicals (see paragraph (n)(8) of this guideline). There exist several variations of the up-and-down experimental design for estimating an LD_{50}. This guideline is derived from the UDP of Bruce as adopted by the American Society for Testing and Materials (ASTM) in 1987 (see paragraph (n)(9) of this guideline) and revised in 1990. A study comparing the results obtained with the UDP, the conventional LD_{50} test and the Fixed Dose Procedure (FDP, OECD Guideline 420) was published in 1995 (see paragraph (n)(10) of this guideline).

(ii) The UDP described in this guideline is of value in minimizing the number of animals required to estimate the acute oral toxicity of a chemical. In addition to the estimation of LD_{50} and CI, the test procedure allows the observation of signs of toxicity. The UDP does not provide information about the slope of the dose-response curve.

(iii) The guideline significantly reduces the number of animals used in comparison to the traditional LD_{50} test, which often required at least 30 animals in a test: (A) The stopping rule limits the number of animals

in a test; (B) sequential dosing introduces further efficiencies in animal use; (C) initial dosing is now set to be below the LD_{50} increasing the percentage of animals in which dosing levels will be sublethal and thereby providing some reduction in pain and distress; and (D) the use of a single sex (usually females) reduces the number of animals needed and minimizes the variability in the test population. In addition, the OECD Guidance Document on Humane Endpoints (see paragraph (n)(11) of this guideline) should be followed in order to reduce the overall suffering of test animals used in this type of toxicity test.

(2) **Initial considerations**—(i) **Choice of starting dose and dose progression factor**. All available information on the test substance should be considered by the testing laboratory prior to conducting the study in order to determine if a preliminary estimate of the LD_{50} and the slope of the dose-response curve can be made. Because the method has a bias toward the starting dose, it is essential that initial dosing occur below the LD_{50}. In addition, the UDP performs best when the spacing between doses or dose progression factor is based on an accurate estimate of the slope of the dose-response curve. (See paragraphs (i)(3)(ii) and (m)(1) of this guideline for discussion of dose sequences and starting values.) Initial information may include the identity and chemical structure of the substance; its physical chemical properties; the results of any other *in vitro* or *in vivo* toxicity tests on the substance or mixtures; toxicological data on structurally related substances or similar mixtures; and the anticipated use(s) of the substance. For example, data from an *in vitro* cytotoxicity assay can also be useful as one of the tools in setting a starting dose for the *in vivo* assessment of acute oral toxicity (see paragraphs (n)(10) through (n)(12) of this guideline). (A Guidance Document on Using *In Vitro* Data to Estimate *In Vivo* Starting Doses for Acute Toxicity is available (see paragraph (n)(11) of this guideline), and preliminary information suggests that the use of this approach may further reduce the number of animals used for *in vivo* testing (see paragraph (n)(11) of this guideline). Preliminary estimates of the LD_{50} and the dose-response slope will help in selecting a dose progression factor and a starting dose for testing.

(ii) **Default starting dose and dose progression factor**. If no information is available to make a preliminary estimate of the LD_{50} and the slope of the dose-response curve, results of computer simulations have suggested that starting near 175 mg/kg and using half-log units (corresponding to a dose progression of 3.2) between doses will produce the best results. This starting dose should be modified if the substance is likely to be highly toxic. The half-log spacing provides for a more efficient use of animals, and increases accuracy in the prediction of the LD_{50} value. However, for chemicals with large variability (i.e., shallow dose- response slopes), bias can still be introduced in the lethality estimates and the LD_{50} estimate will have a large statistical error, similar to other acute toxicity methods. To correct for this, the main test includes a stopping rule keyed

to properties of the estimate rather than a fixed number of test observations. (See paragraph (i)(3)(iii) of this guideline.)

(iii) **Delayed toxicity**. The method is easiest to apply to materials that produce death within one or two days. The method would not be practical to use when considerably delayed death (five days or more) can be expected.

(iv) **Computation**. Computers are used to facilitate animal-by-animal calculations that establish testing sequences and provide final estimates. The users of this protocol are strongly urged to solely use the Agency-developed software package (AOT425StatPgm) for performing the test and the calculation of the LD 50. The software is available on EPA's Internet Web site at http://www.epa.gov/oppfead1/harmonized.

(v) **Humane practices**. Moribund animals or animals obviously in pain or showing signs of severe and enduring distress shall be humanely killed, and are considered in the interpretation of the test results in the same way as animals that died on test. Criteria for making the decision to kill moribund or severely suffering animals, and guidance on the recognition of predictable or impending death are the subject of an OECD guidance document (see paragraph (n)(11) of this guideline).

(vi) **Limit test**. A limit test can be used efficiently to identify chemicals that are likely to have low acute toxicity.

(f) **Principle of the limit test**. The limit test is a sequential test that uses a maximum of 5 animals (see paragraphs (i)(2)(i) through (i)(2)(iv) of this guideline). A test dose of 5000 mg/kg is used. The selection of a sequential test plan increases the statistical power and also has been made to intentionally bias the procedure towards rejection of the limit test for compounds with LD_{50}s near the limit dose; i.e., to err on the side of safety. As with any limit test protocol, the probability of correctly classifying a compound will decrease as the actual LD_{50} more nearly resembles the limit dose.

(g) **Principle of the Main Test**. (1) The main test consists of a single ordered dose progression in which animals are dosed, one at a time, at 48-hour intervals. The first animal receives a dose a step below the level of the best estimate of the LD_{50}. If the animal survives, the dose for the next animal is increased to a factor of one half log times the original dose; if it dies, the dose for the next animal is decreased by a similar dose progression. (Note: 3.2 is the default factor corresponding to a dose progression of one half log unit in the Agency developed software program (AOT425StatPgm). However, this value may be changed. Paragraphs (i)(3)(ii) and (m)(12) of this guideline provide further guidance for choice of dose spacing factor.) Each animal should be observed carefully for up to 48 hours before making a decision on whether and how much to dose the next animal. That decision is based on the 48-hour survival pattern

of all the animals up to that time. (See paragraphs (i)(3)(i) and (i)(3)(v) of this guideline on choice of survival interval.) A combination of stopping criteria is used to keep the number of animals low while adjusting the dosing pattern to reduce the effect of a poor starting value or low slope (see paragraph (i)(3)(iv) of this guideline). Dosing is stopped when one of these criteria is satisfied (see paragraphs (i)(3)(iii) and (k)(2) of this guideline), at which time an estimate of the LD_{50} and a CI are calculated for the test based on the status of all the animals at termination. For most applications, testing will be completed with only 4 animals after initial reversal in animal outcome. The LD_{50} is calculated using the method of maximum likelihood (see paragraphs (k)(2) and (k)(2)(iii) of this guideline.)

(2) The results of the main test procedure serve as the starting point for a computational procedure to provide a CI estimate where feasible. A description of the basis for this CI is outlined in paragraph (k)(3) of this guideline.

(h) **Preparation for testing**—(1) **Selection of animals species.** The preferred rodent species is the rat although other rodent species may be used.

(2) **Single sex selection**. The test is conducted using a single sex in order to reduce variability and as a means of minimizing the number of animals used. Either sex may be used, however, if there is information available indicating differences in sensitivity, the most sensitive sex (usually females) should be tested (see paragraph (n)(11) of this guideline).

(i) Literature surveys of conventional LD_{50} tests show that usually there is little difference in sensitivity between the sexes but, in those cases where differences were observed, females were often slightly more sensitive (see paragraph (n)(10) of this guideline). For chemicals that are direct acting in their toxic mechanism, female rats may have a lower detoxification capacity than males, as measured by specific activity of phase I and II enzymes. However, all available information should be evaluated, for example on chemical analogues and the results of testing for other toxicological endpoints on the chemical itself, as this may indicate that males may be more sensitive than females. Knowledge that metabolic activation is required for a chemical's toxicity can also indicate that males may be the more sensitive sex.

(ii) Occasionally, the results of subsequent testing, for example a subchronic test, may raise concerns that the more sensitive sex had not been used. In such cases, and only when considerable differences between the sexes are suspected, it may be necessary to conduct another full acute oral toxicity study in the second sex. This is preferable to conducting confirmatory testing in a small group of animals of the second sex as a late satellite to the original test because there is a strong possibility that this

would produce results that are difficult to interpret. The impact of conducting a second full test on the overall number of animals used in acute toxicity testing should be small because re-testing is anticipated to be infrequent and the results of the test in one sex, together with data from any subsequent studies, will greatly assist in the selection of starting doses closer to the LD_{50} in the second test.

(3) **Age and weight ranges**. Healthy young adult animals of commonly used laboratory strains should be employed. Females should be nulliparous and non-pregnant. At the commencement of its dosing, each animal should be between 8 weeks and 12 weeks old. In order to minimize the contribution of developmental variability to study outcome, 10 weeks, with a range of ± 1 week is recommended if practical. The weight of each animal should fall in an interval ± 20% of the mean initial weight of all previously dosed animals.

(4) **Housing and feeding conditions**. The temperature in the experimental animal room should be 22°C (± 3°C). The relative humidity should be at least 30% and preferably not exceed 70% other than during room cleaning. Lighting should be artificial, the sequence being 12 hours light and 12 hours dark. The animals are housed individually. For feeding, conventional rodent laboratory diets may be used with an unlimited supply of drinking water.

(5) **Preparation of animals**. The animals are randomly selected, marked to permit individual identification, and kept in their cages for at least 5 days prior to dosing to allow for acclimatization to the laboratory conditions. As with other sequential test designs, care must be taken to ensure that animals are available in the appropriate size and age range for the entire study.

(6) **Preparation of doses**. (i) When necessary, the test substance is dissolved or suspended in a suitable vehicle. The use of an aqueous solution/suspension/emulsion is recommended wherever possible, followed in order of preference by a solution/suspension/emulsion in oil (e.g. corn oil) and then possibly solution in other vehicles. For vehicles other than water the toxicological characteristics of the vehicle should be known. Dosing preparations must be prepared shortly prior to administration unless the stability of the preparation over the period during which it will be used is known. Where preparation shortly before administration is not practicable and the stability of the preparation is not known, this will need to be demonstrated analytically.

(ii) Constant concentration should be used in dosing unless there is clear scientific or regulatory justification for not doing so. The maximum dose volume for administration must not be exceeded. The maximum volume of liquid that can be administered at one time depends on the size of the test animal. In rodents, the volume should not normally exceed

1 ml/100g of body weight; however, in the case of aqueous solutions, 2 ml/100g body weight can be considered.

(7) **Administration of doses.** (i) The test substance is administered in a single dose by gavage using a stomach tube or a suitable intubation cannula. In the unusual circumstance that a single dose is not possible, the dose may be given in smaller fractions over a period not exceeding 24 hours.

(ii) Animals should be fasted prior to dosing (e.g., with the rat, food but not water should be withheld overnight; with the mouse, food but not water should be withheld for 3–4 hours). Following the period of fasting, the animals should be weighed and the test substance administered. The fasted body weight of each animal is determined and the dose is calculated according to the body weight. After the substance has been administered, food may be withheld for a further 3–4 hours in rats or 1–2 hours in mice. Where a dose is administered in fractions over a period of time, it may be necessary to provide the animals with food and water depending on the length of the period.

(i) **The up-and-down testing procedure**—(1) **Choice of limit test and main test.** The limit test is primarily used in situations where the experimenter has information indicating that the test material is likely to be nontoxic, i.e., having toxicity below regulatory limit doses. Information about the toxicity of the test material can be gained from knowledge about similar tested compounds or similar tested mixtures or products, taking into consideration the identity and percentage of components known to be of toxicological significance. In those situations where there is little or no information about its toxicity, or in which the test material is expected to be toxic, the main test should be performed.

(2) **Implementation of the limit test.** (i) The Agency has developed dedicated software for performing the test and calculation of test results (see paragraph (e) (2)(iv) of this guideline).

(ii) Dose one animal at 5000 mg/kg. If the animal dies, conduct the main test starting at 175 mg/kg to determine the LD_{50}. If the animal survives, dose two additional animals. If both animals survive, the LD_{50} is greater than the limit dose and the test is terminated (i.e. carried to full 14-day observation without dosing of further animals). If one or both animals die, then dose an additional two animals, one at a time. If an animal unexpectedly dies late in the study, and there are other survivors, it is appropriate to stop dosing and observe all animals to see if other animals will also die during a similar observation period (see paragraph (g)(1) of this guideline for initial observation period). Late deaths should be counted the same as other deaths. The results are evaluated as follows (O=survival and X=death).

(iii) The LD_{50} is less than the test dose (5000 mg/kg) when three or more animals die. If a third animal dies, conduct the main test.

O XO XX

O OX XX

O XX OX

O XX X

(iv) The LD_{50} is greater than the test dose (5000 mg/kg) when three or more animals survive.

O OO

O XO XO

O XO O

O OX XO

O OX O

O XX OO

(v) If a limit test is performed at 2000 mg/kg, animals should be dosed sequentially and testing should be performed on all five animals.

(3) **Implementation of the main test.** (i) The Agency has developed dedicated software for performing the test and calculation of test results (see paragraph (e) (2)(iv) of this guideline).

(ii) Performing the UDP. Single animals are dosed in sequence usually at 48-hour intervals. However, the time interval between dosing is determined by the onset, duration, and severity of toxic signs. Treatment of an animal at the next dose should be delayed until one is confident of survival of the previously dosed animal. The time interval may be adjusted as appropriate, e.g., in case of inconclusive response. The test is simpler to implement when a single time interval is used for making sequential dosing decisions. Nevertheless, it is not necessary to recalculate dosing or likelihood-ratios if the time interval changes midtest. For selecting the starting dose, all available information, including information on structurally related substances and results of any other toxicity tests on the test material, should be used to approximate the LD_{50} as well as the slope of the dose-response curve.

(iii) Choice of starting dose and dose progression. The first animal is dosed a step below the toxicologist's best estimate of the LD_{50}. If the animal survives, the second animal receives a higher dose. If the first animal dies or appears moribund, the second animal receives a lower dose. The same dosing decision pattern is followed for each subsequent animal.

The dose progression factor should be chosen to be the antilog of 1/(the estimated slope of the dose-response curve) (a progression of 3.2 corresponds to a slope of 2) and should remain constant throughout testing. Thus, when there is no information on the slope of the substance to be tested, a default dose progression factor of 3.2 is used. Using the default progression factor, doses would be selected from the sequence 1.75, 5.5, 17.5, 55, 175, 550, 1750, 5000. If no estimate of the substance's lethality is available, dosing should be initiated at 175 mg/kg. In most cases, this dose is sublethal and therefore serves to reduce the level of pain and suffering. If animal tolerances to the chemical are expected to be highly variable (i.e., slopes are expected to be less than 2.0), consideration should be given to increasing the dose progression factor beyond the default 0.5 on a log dose scale (i.e., 3.2 progression factor) prior to starting the test. Similarly, for test substances known to have very steep slopes, dose progression factors smaller than the default should be chosen. (Paragraph (m)(3) of this guideline relates choice of dose progression to assumed slope and *sigma* and discusses test performance. Paragraph (m)(1) of this guideline includes a table of dose progressions for whole number slopes ranging from 1 to 8 with starting dose 175 mg/kg.)

(iv) Stopping rules. Dosing continues depending on the fixed-time interval (e.g., 48-hours) outcomes of all the animals up to that time. The testing stops when one of the following stopping criteria first is met:

(A) 3 consecutive animals survive at the upper bound;

(B) 5 reversals occur in any 6 consecutive animals tested;

(C) At least 4 animals have followed the first reversal and the specified likelihood-ratios exceed the critical value. (See paragraphs (k)(2)(iv) and (m)(2) of this guideline). Calculations are made at each dosing, following the fourth animal after the first reversal.).

(v) Total number of doses. For a wide variety of combinations of LD_{50} and slopes, stopping rule in paragraph (i)(3)(iii)(C) of this guideline will be satisfied with 4 to 6 animals after the test reversal. In some cases for chemicals with shallow slope dose-response curves, additional animals (up to a total of fifteen tested) may be needed.

(vi) Calculation. When the stopping criteria have been attained, the estimated LD_{50} should be calculated from the animal outcomes at test termination using the method described in paragraphs (k)(1)(i) and (k)(2)(i) of this guideline.

(vii) Humane practices. Moribund animals killed for humane reasons are considered in the same way as animals that died on test. If an animal unexpectedly dies late in the study and there are other survivors at that dose or above, it is appropriate to stop dosing and observe all animals to see if other animals will also die during a similar observation period.

If subsequent survivors also die, and it appears that all dose levels exceed the LD_{50} it would be most appropriate to start the study again beginning at least two steps below the lowest dose with deaths (and increasing the observation period) since the technique is most accurate when the starting dose is below the LD_{50}. If subsequent animals survive at or above the dose of the animal that dies, it is not necessary to change the dose progression since the information from the animal that has now died will be included into the calculations as a death at a lower dose than subsequent survivors, pulling the LD_{50} down.

(j) **Observations**. Animals are observed individually at least once during the first 30 minutes after dosing, periodically during the first 24 hours (with special attention given during the first 4 hours), and daily thereafter, for a total of 14 days, except where they need to be removed from the study and humanely killed for animal welfare reasons or are found dead. However, the duration of observation should not be fixed rigidly. It should be determined by the toxic reactions and time of onset and length of recovery period, and may thus be extended when considered necessary. The times at which signs of toxicity appear and disappear are important, especially if there is a tendency for toxic signs to be delayed (see paragraph (n)(15) of this guideline). All observations of toxic signs are systematically recorded with individual records being maintained for each animal. Additional observations will be necessary if the animals continue to display signs of toxicity.

(1) **Toxic signs**. Observations should include changes in skin and fur, eyes and mucous membranes, and also respiratory, circulatory, autonomic and central nervous systems, and somatomotor activity and behavior pattern. Attention should be directed to observations of tremors, convulsions, salivation, diarrhea, lethargy, sleep and coma. The principles and criteria summarized in the Humane Endpoints Guidance Document (see paragraph (n)(11) of this guideline) should be taken into consideration. Animals found in a moribund condition and animals showing severe pain and enduring signs of severe distress should be humanely killed. When animals are killed for humane reasons or found dead, the time of death should be recorded as precisely as possible.

(2) **Body weight**. Individual weights of animals should be determined shortly before the test substance is administered and at least weekly thereafter. Weight changes should be calculated and recorded. At the end of the test surviving animals are weighed and then humanely killed.

(3) **Pathology**. All animals (including those which die during the test or are removed from the study for animal welfare reasons) should be subjected to gross necropsy. All gross pathological changes should be recorded for each animal. Microscopic examination of organs showing evidence of gross pathology in animals surviving 24 or more hours after the

initial dosing may also be considered because it may yield useful information.

(k) **Data and reporting**—(1) **Data**. Individual animal data should be provided. Additionally, all data should be summarized in tabular form, showing for each test dose the number of animals used, the number of animals displaying signs of toxicity (see paragraph (n)(15) of this guideline), the number of animals found dead during the test or killed for humane reasons, time of death of individual animals, a description and the time course of toxic effects and reversibility, and necropsy findings. A rationale for the starting dose and the dose progression and any data used to support this choice should be provided.

(2) **Calculation of LD_{50} for the main test**—(i) **Maximum likelihood**. The LD_{50} is calculated using the maximum likelihood method, except in the exceptional cases described in paragraphs (k)(2)(ii) and (m)(3) of this guideline. The Agency-developed software program (AOT425StatPgm) available on EPA's Internet Web site at http://www.epa.gov/oppfead1/harmonized should be used to perform this calculation. The following statistical details may be helpful in implementing the maximum likelihood calculations suggested (with an assumed *sigma*). All deaths, whether immediate or delayed or humane kills, are incorporated for the purpose of the maximum likelihood analysis. Following Dixon (see paragraph (n)(5) of this guideline), the likelihood function is written as follows:

$L = L_1 L_2 \ldots . L_n$,

where

L is the likelihood of the experimental outcome, given μ and *sigma*, and n the total number of animals tested.

$L_i = 1 - F(Z_i)$ if the i^{th} animal survived, or

$L_i = F(Z_i)$ if the i^{th} animal died,

where

F = cumulative standard normal distribution,

$Z_i = [\log(d_i) - \mu\ ] / sigma$

d_i = dose given to the i^{th} animal, and

sigma = standard deviation in log units of dose (which is not the log standard deviation).

An estimate of the log of the true LD_{50} is given by the value of μ that maximizes the likelihood L (see paragraph (k)(2)(iii) of this guideline).

An estimate of *sigma* of 0.5 is used unless a better generic or case-specific value is available.

(ii) **Special circumstances**. Under some circumstances, statistical computation will not be possible or will likely give erroneous results. Special means to determine/report an estimated LD_{50} are available for these circumstances as described in the following paragraphs (k)(2)(ii)(A), (k)(2)(ii)(B), and (k)(2)(ii)(C). If none of these situations occurs, then the LD_{50} is calculated using the maximum likelihood method.

(A) If testing stopped based on the criterion in paragraph (i)(3)(iii)(C) of this guideline (i.e., a boundary dose was tested repeatedly), or if the upper bound dose ended testing, then the LD_{50} is reported to be above the upper bound.

(B) If all the dead animals have higher doses than all the live animals (or if all live animals have higher doses than all the dead animals, although this is practically unlikely), then the LD_{50} is between the doses for the live and the dead animals. These observations give no further information on the exact value of the LD_{50}. Still, a maximum likelihood LD_{50} estimate can be made provided there is a prior value for *sigma*. The LD_{50} estimate is only as good as the validity of the assumed signa. However, Case 3 as described in paragraph (m)(3)(iii) of this guideline and here is most likely to occur because the dose progression (based on the assumed signma) is too wide. The stopping criterion in paragraph (i)(3)(iii)(C) describes one such circumstance.

(C) If the live and dead animals have only one dose in common and all the other dead animals have higher doses and all the other live animals lower doses, or vice versa, then the LD_{50} equals their common dose. If a closely related substance is tested, testing should proceed with a smaller dose progression.

(iii) **Maximum likelihood calculation**. Maximum likelihood calculation should be performed using a dedicated program developed by and available from EPA (see paragraph (n)(16) of this guideline). If other computer programs are used, the laboratory should take care in handling special cases described in this guideline and the documentation of test performance available on EPA's Internet Web site at http://www.epa.gov/oppfead1/harmonized. Typical instructions for these packages are given in appendices to the ASTM Standard E 1163-87 (see paragraph (n)(9) of this guideline). (The *sigma* used in the BASIC program in (see paragraph (n)(9) of this guideline) will need to be edited to reflect the parameters of the UDP.) The program's output is an estimate of log (LD_{50}) and its standard error.

(iv) **Stopping rule**. The likelihood-ratio stopping rule in paragraph (i)(3)(iii)(C) of this guideline is based on three measures of test progress, that are of the form of the likelihood in paragraph (k)(2) of this guideline,

with different values for μ. Comparisons are made after each animal tested after the sixth that does not already satisfy the criteria in paragraph (i)(3)(iii)(A) or paragraph (i)(3)(iii)(B) guideline. The equations for the likelihood-ratio criteria are provided by following the steps in paragraph (m)(2)(vii) of this guideline. These comparisons are most readily performed in an automated manner and can be executed repeatedly, for instance, by a spreadsheet routine such as that also provided in paragraph (m)(2)(vii) of this guideline. If the criterion is met, testing stops and the LD_{50} can be calculated by the maximum likelihood method.

(3) **Computation of CI.** (i) Following the main test and estimated LD_{50} calculation, it may be possible to compute interval estimates for the LD_{50}. The Agency-developed software program AOT425StatPgm will perform the calculations. Any of these CIs provides valuable information on the reliability and utility of the main test that was conducted. A wide CI indicates that there is more uncertainty associated with the estimated LD_{50}. In this case, the reliability of the estimated LD_{50} is low and the usefulness of the estimated LD_{50} may be marginal. A narrow interval indicates that there is relatively little uncertainty associated with the estimated LD_{50}. In this case, the reliability of the estimated LD_{50} is high and the usefulness of the estimated LD_{50} is good. This means that if the main test were to be repeated, the new estimated LD_{50} is expected to be close to the original estimated LD_{50} and both of these estimates are expected to be close to the true LD_{50}.

(ii) Depending on the outcome of the main test, one of two different types of interval estimates of the true LD_{50} is calculated:

(A) When at least three different doses have been tested and the middle dose has at least one animal that survived and one animal that died, a profile-likelihood-based computational procedure is used to obtain a CI that is expected to contain the true LD_{50} 95% of the time. However, because small numbers of animals are expected to be used, the actual level of confidence is generally not exact (see paragraph (n)(19) of this guideline). The random stopping rule improves the ability of the test overall to respond to varying underlying conditions, but also causes the reported level of confidence and the actual level of confidence to differ somewhat (see paragraph (n)(18) of this guideline).

(B) If all animals survive at or below a given dose level and all animals die when dosed at the next higher dose level, an interval is calculated that has as its lower limit the highest dose tested where all the animals survive and has as its upper limit the dose level where all the animals died. This interval is labeled as "approximate." The exact confidence level associated with this interval cannot be specifically determined. However, because this type of response would only occur when the dose-response is steep, in most cases, the true LD_{50} is expected to be contained

within the calculated interval or be very close to it. This interval will be relatively narrow and sufficiently accurate for most practical use.

(iii) In some instances, CIs are reported as infinite, through including either zero at the lower end or infinity at the upper end, or both. Such intervals may occur, for example, when the response profile is relatively flat or relatively uncertain.

(iv) Implementing this set of procedures requires specialized computation which is either by use of a dedicated program to be available through the Environmental Protection Agency (EPA) or OECD or developed following technical details available from the EPA or OECD. Achieved coverage of these intervals and properties of the dedicated program are described in a report (see paragraph (n)(16) of this guideline) also available through the EPA. Paragraph (m)(3) of this guideline provides information on choice of dose progression and initial dose level for the UDP and describes test performance under a variety of circumstances.

(l) **Test reporting**. The test report must include the following information:

(1) Test substance:

(i) Physical nature, purity and physicochemical properties (including isomerization);

(ii) Identification data.

(2) Vehicle (if appropriate): Justification for choice of vehicle, if other than water.

(3) Test animals:

(i) Species/strain used;

(ii) Microbiological status of the animals, when known;

(iii) Number, age and sex of animals;

(iv) Rationale for use of males instead of females;

(v) Source, housing conditions, diet, etc.;

(vi) Individual weights of animals at the start of the test, at day 7, and at day 14.

(4) Test conditions:

(i) Rationale for initial dose level selection, dose progression factor and for follow-up dose levels;

(ii) Details of test substance formulation;

(iii) Details of the administration of the test substance;

(iv) Details of food and water quality (including diet type/source, water source).

(5) Results:

(i) Body weight/body weight changes;

(ii) Tabulation of response data by sex (if both sexes are used) and dose level for each animal (i.e., animals showing signs of toxicity including nature, severity, duration of effects, and mortality);

(iii) Time course of onset of signs of toxicity and whether these were reversible for each animal;

(iv) Necropsy findings and any histopathological findings for each animal, if available;

(v) LD_{50} and CI (which the AOT425StatPgm software package uses);

(vi) Statistical treatment of results (description of computer routine used and spreadsheet tabulation of calculations). If other than Agency-supplied software is used, give explanation of now the program was verified against Agency software.

(6) Discussion and interpretation of results.

(7) Conclusions.

(m) **Additional guidance for toxicologists**—(1) **Dosing procedure—dose sequence for main test.** (i) Up-and-down dosing procedure. For each run, animals are dosed, one at a time, usually at 48-hour intervals. The first animal receives a dose a step below the level of the best estimate of the LD_{50}. This selection reflects an adjustment for a tendency to bias away from the LD_{50} in the direction of the initial starting dose in the final estimate (see paragraph (e)(2)(ii) of the guideline). The overall pattern of outcomes is expected to stabilize as dosing is adjusted for each subsequent animal. Paragraph (m)(1)(iii) of this guideline provides further guidance for choice of dose spacing factor.

(ii) Default dose progression. Once the starting dose and dose spacing are decided, the toxicologist should list all possible doses including the upper bound (usually 2000 or 5000 mg/kg). Doses that are close to the upper bound should be removed from the progression. The stepped nature of the UDP design provides for the first few doses to function as a self-adjusting sequence. Because of the tendency for positive bias, in the event that nothing is known about the substance, a starting dose of 175 mg/kg is recommended. If the default procedure is to be used for the main test, dosing will be initiated at 175 mg/kg and doses will be spaced by a factor of 0.5 on a log dose scale. The doses to be used include 1.75,

5.5, 17.5, 55, 175, 550, 2000 or, for specific regulatory needs, 1.75, 5.5, 17.5, 55, 175, 550, 1750, 5000. For certain highly toxic substances, the dosing sequence may need to be extended to lower values.

(iii) In the event a dose progression factor other than the default is deemed suitable, the following Table 1 provides dose progressions for whole number multiples of slope, from 1 to 8. (See paragraph (m)(3) of this guideline for discussion of influence of dose progression on test performance.)

Table 1.—Dose Progressions for UDP

(Choose a Slope and Read Down the Column. All doses in mg/kg body weight)

Slope =	1	2	3	4	5	6	7	8
	0.175*	0.175*	0.175*	0.175*	0.175*	0.175*	0.175*	0.175*
							0.243*	0.233*
					0.28	0.26		
				0.31			0.34	0.31
			0.38			0.38		
								0.41
					0.44		0.47	
		0.55		.55		0.55		0.55
				0.70		0.65		
							0.74	
			.81			.81		
				0.98			0.91	0.98
					110	1.19		
							1.26	1.31
	1.75	1.75	1.75	1.75	1.75	1.75	1.75	1.75
							2.43	2.33
					2.8	2.6		
				3.1			3.4	3.1
			3.8			3.8		
					4.4			4.1
							4.7	
		5.5		5.5	5.5		5.5	
					7.0		6.5	
								7.4
			8.1			8.1		
				9.8			9.1	9.8
					11.0	11.9		
							12.6	13.1
	17.5	17.5	17.5	17.5	17.5	17.5	17.5	17.5
							24.3	23.3
					28	26		
				31			34	31
			38			38		
					44			41
							47	
		55		55		55		55
							65	
					70			74
			81			81		
				98			91	98
					110	119		
							126	131
	175	175	175	175	175	175	175	175
							243	233
					280	260		
				310			340	310
			380			380		
					440			410
							470	
		550		550		550		550
							650	
					700			740
			810			810		
				980			910	980
					1100	1190		
							1260	1310
	1750	1750	1750	1750	1750	1750	1750	1750
							2430	2330
					2800	2600		
				3100				3100
						3800	3400	
								4100
	5000	5000	5000	5000	5000	5000	5000	5000

* If lower doses are needed, continue progressions to a lower dose

(2) **Computations for the likelihood-ratio stopping rules.** (i) As described in paragraph (i)(3)(iii) of this guideline, the main test may be completed on the basis of the first of three stopping criteria to occur. In any case, even if none of the stopping criteria is satisfied, dosing would stop when 15 animals are dosed. Tables 2, 4, and 6 in paragraphs (m)(2)(ii), (m)(2)(iii), and (m)(2)(iv), respectively, of this guideline illustrate examples where testing has started with no information, so the rec-

ommended default starting value, 175 mg/kg, and the recommended default dose progression factor, 3.2 or one half log, have been used. Tables 3, 5, and 7 in paragraphs (m)(2)(ii), (m)(2)(iii), and (m)(2)(iv), respectively, illustrate how Tables 2, 4, and 6, respectively, would appear in the dedicated program referenced in paragraph (k)(3)(iv) (see also paragraph (n)(16)).

(ii) The following Tables 2 and 3 show how the main test would stop if 3 animals have survived at the limit dose of 5000 mg/kg. (This example illustrates situations where a limit test was not thought appropriate *a priori*).

Table 2. Example of Stopping Criterion in Paragraph (i)(3)(iii)(A) using 5000 mg/kg.

Stop after animal #6 because 3 animals survive at limit of 5000 mg/kg (#4-#6).

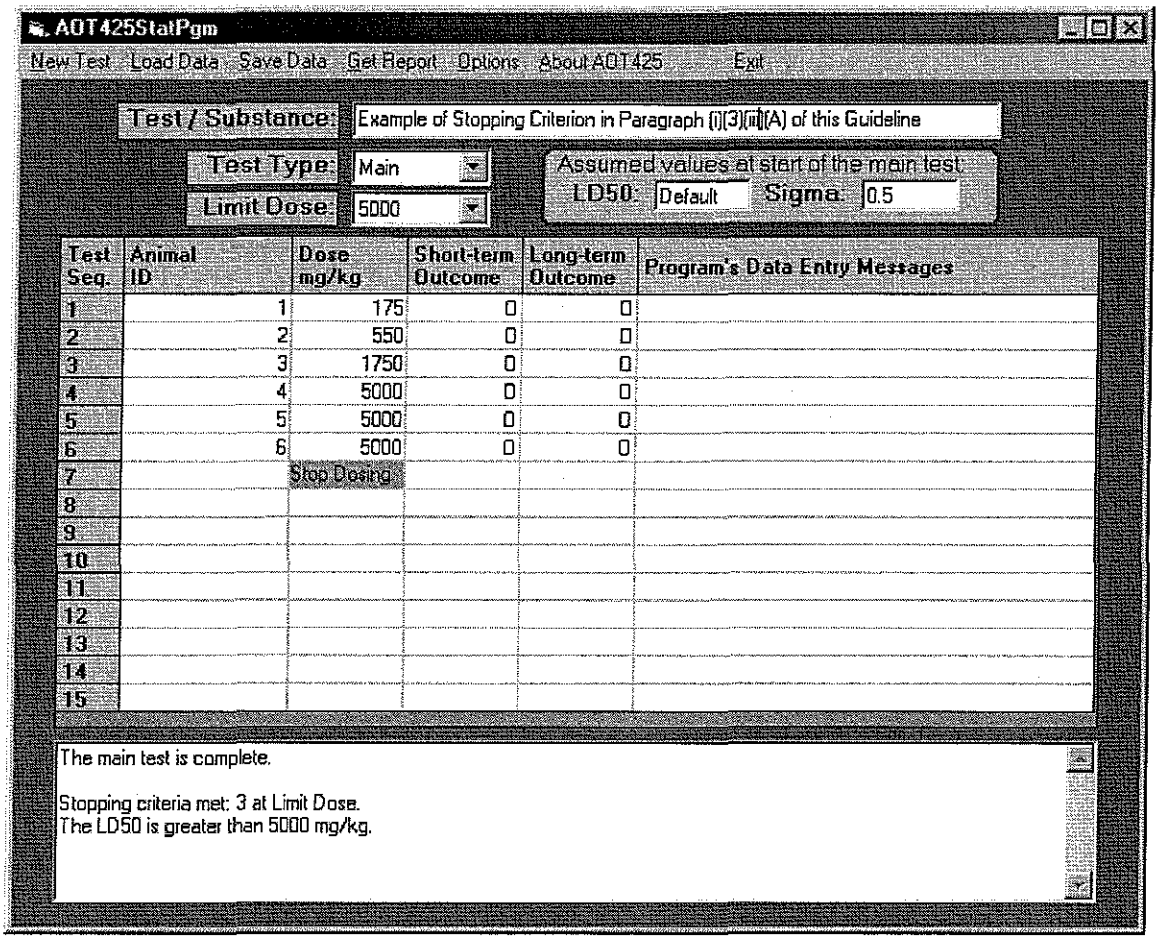
1	2	3	4	5	6	7	8	9	10	11	12
Step	(I)nclude; (E)xclude	Dose	(X)response (O)non-resp. OK	Included in nominal *n*	log10 Dose	LD50 = Prob. of response	#DIV/0! likelihood contribn. (ln *Li*)	LD50 = Prob. of response	#DIV/0! likelihood contribn. (ln *Li*)	LD50 = Prob. of response	#DIV/0! likelihood contribn. (ln *Li*)
1	I	175	O	no	2.2430	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
2	I	550	O	no	2.7404	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
3	I	1750	O	no	3.2430	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
4	I	5000	O	no	3.6990	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
5	I	5000	O	no	3.6990	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
6	I	5000	O	no	3.6990	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
7	E				-						
8	E				-						
9	E				-	-	-	-	-	-	-
10	E				-	-	-	-	-	-	-
11	E				-	-	-	-	-	-	-
12	E				-				-	-	-
13	E				-				-	-	-
14	E				-				-	-	-
15	E				-	-	-	-	-	-	-
Nominal Sample size =				0							
Actual number tested =				6							

Ignore all calculation cells. No reversal in direction of response.

Maximum Likelihood Calculations cannot be completed. LD50 is greater than 5000 mg/kg.

Calculated maximum likelihood estimate of LD50 = none

Table 3. Example of Stopping Criterion in Paragraph (i)(3)(iii)(A) of this Guideline Using 5000 mg/kg



(iii) The following Tables 4 and 5 show how a particular sequence of 5 reversals in 6 tested animals could occur and allow test completion.

Table 4. Example of Stopping Criterion in Paragraph (i)(3)(iii)(B).

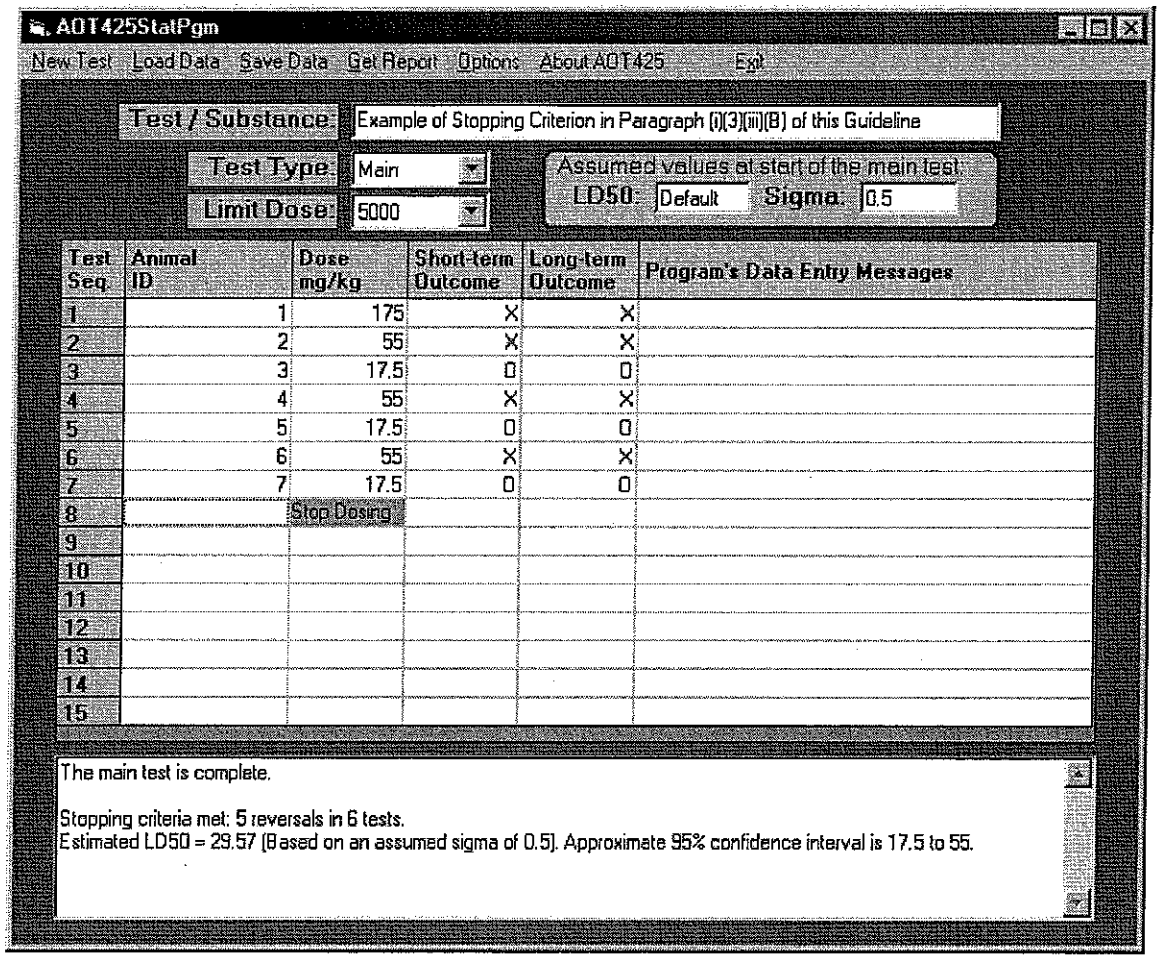
Stop after animal #7 because 5 reversals in 6 consecutive animals tested (#2-#7).

1	2	3	4	5	6	7	8	9	10	11	12
Step	(I)nclude; (E)xclude	Dose	(X)response (O)non-resp. OK	Included in nominal *n*	log10 Dose	LD50 = Prob. of response	31.0 likelihood contribn. (ln *Li*)	LD50 = Prob. of response	12.4 likelihood contribn. (ln *Li*)	LD50 = Prob. of response	77.6 likelihood contribn. (ln *Li*)
1	I	175	X	no	2.2430	0.9335	-0.0688	0.9892	-0.0108	0.7602	-0.2742
2	I	55	X	yes	1.7404	0.6905	-0.3703	0.9020	-0.1031	0.3826	-0.9607
3	I	17.5	O	yes	1.2430	0.3095	-0.3703	0.6174	-0.9607	0.0980	-0.1031
4	I	55	X	yes	1.7404	0.6905	-0.3703	0.9020	-0.1031	0.3826	-0.9607
5	I	17.5	O	yes	1.2430	0.3095	-0.3703	0.6174	-0.9607	0.0980	-0.1031
6	I	55	X	yes	1.7404	0.6905	-0.3703	0.9020	-0.1031	0.3826	-0.9607
7	I	17.5	O	yes	1.2430	0.3095	-0.3703	0.6174	-0.9607	0.0980	-0.1031
8	E				-	-	-	-	-	-	-
9	E				-	-	-	-	-	-	-
10	E				-	-	-	-	-	-	-
11	E				-	-	-	-	-	-	-
12	E				-	-	-	-	-	-	-
13	E				-	-	-	-	-	-	-
14	E				-	-	-	-	-	-	-
15	E				-	-	-	-	-	-	-
Nominal Sample size =				6							
Actual number tested =				7							
Dose-averaging estimator				31.02							
log10 =				1.492							
log-likelihood sums:							-2.2906		-3.2021		-3.4655
likelihoods:							0.1012		0.0407		0.0313
likelihood ratios:									2.4880		3.2378
Individual ratios exceed critical value?				critical=	2.5				FALSE		TRUE
Both ratios exceed critical value?									FALSE		
Calculated maximum likelihood estimate of LD50 =					29.6						

Automated calculation; not relevant to this case.

Final estimate obtained from Maximum Likelihood Calculations

Table 5. Example of Stopping Criterion in Paragraph (i)(3)(iii)(B) of this Guideline.



(iv) Finally, the following Tables 6 and 7 illustrate a situation several animals into a test, where neither the criterion in paragraph (i)(3)(iii)(A) nor the criterion in paragraph (i)(3)(iii)(B) of this guideline has been met, a reversal of response has occurred followed by 4 tested animals, and, consequently, the criterion in paragraph (i)(3)(iii)(C) of this guideline must be evaluated as well.

Table 6. Example of Stopping Criterion in Paragraph (i)(3)(iii)(C).

Stop when LR criterion is first met, here at animal #9.
Check LR criterion starting at animal #6.

Assumed slope	2	sigma =	0.5

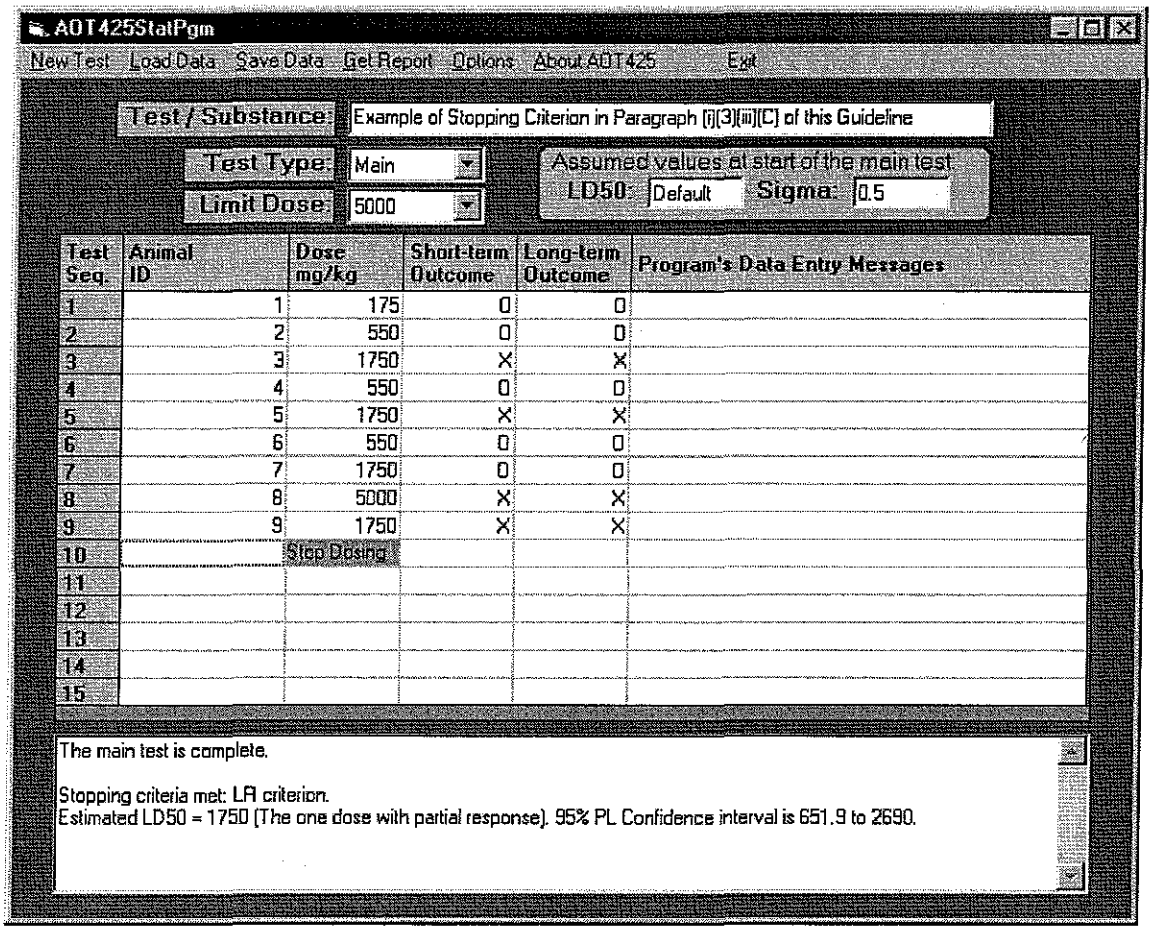
Parameters of convergence criterion	
critical LR	2.5
factor of LD50	2.5

Result: The LR criterion is met

1	2	3	4	5	6		7	8	9	10	11	12
Step	(I)nclude; (E)xclude	Dose	(X)response (O)non-resp. OK	Included in nominal *n*	log10 Dose	Contrib.to DAE	LD50 = Prob. of response	1292.8 likelihood contribn. (ln *Li*)	LD50 = Prob. of response	517.1 likelihood contribn. (ln *Li*)	LD50 = Prob. of response	3232.0 likelihood contribn. (ln *Li*)
1	I	175	O	no	2.2430	0.0000	0.0412	-0.0421	0.1733	-0.1903	0.0057	-0.0057
2	I	550	O	yes	2.7404	2.7404	0.2289	-0.2600	0.5214	-0.7368	0.0620	-0.0640
3	I	1750	X	yes	3.2430	3.2430	0.6037	-0.5046	0.8552	-0.1564	0.2971	-1.2138
4	I	550	O	yes	2.7404	2.7404	0.2289	-0.2600	0.5214	-0.7368	0.0620	-0.0640
5	I	1750	X	yes	3.2430	3.2430	0.6037	-0.5046	0.8552	-0.1564	0.2971	-1.2138
6	I	550	O	yes	2.7404	2.7404	0.2289	-0.2600	0.5214	-0.7368	0.0620	-0.0640
7	I	1750	O	yes	3.2430	3.2430	0.6037	-0.9257	0.8552	-1.9323	0.2971	-0.3525
8	I	5000	X	yes	3.6990	3.6990	0.8800	-0.1279	0.9756	-0.0247	0.6477	-0.4344
9	I	1750	X	yes	3.2430	3.2430	0.6037	-0.5046	0.8552	-0.1564	0.2971	-1.2138
10	E				-	0.0000	-	-	-	-	-	-
11	E				-	0.0000	-	-	-	-	-	-
12	E				-	0.0000	-	-	-	-	-	-
13	E				-	0.0000	-	-	-	-	-	-
14	E				-	0.0000	-	-	-	-	-	-
15	E				-	0.0000	-	-	-	-	-	-
Nominal Sample size =				8								
Actual number tested =				9								
Dose-averaging estimator				1292.78								
log10 =				3.112								
log-likelihood sums:								-3.3894		-4.8270		-4.6260
likelihoods:								0.0337		0.0080		0.0098
likelihood ratios:										4.2104		3.4436
Individual ratios exceed critical value?				critical=	2.5					TRUE		TRUE
Both ratios exceed critical value?										TRUE		
Calculated maximum likelihood estimate of LD50 =						1329.6						

Final estimate obtained from Maximum Likelihood Calculations

Table 7. Example of Stopping Criterion in Paragraph (i)(3)(iii)(C) of this Guideline.



v

(v) Criterion in paragraph (i)(3)(iii)(C) of this guideline calls for a likelihood-ratio stopping rule to be evaluated after testing each animal, starting with the fourth tested following the reversal. Three ''measures of test progress'' are calculated. Technically, these measures of progress are likelihoods, as recommended for the maximum-likelihood estimation of the LD_{50}. The procedure is closely related to calculation of a CI by a likelihood-based procedure.

(vi) The basis of the procedure is that when enough data have been collected, a point estimate of the LD_{50} should be more strongly supported than values above and below the point estimate, where statistical support is quantified using likelihood. Therefore three likelihood values are calculated: A likelihood at an LD_{50} point estimate (called the rough estimate or dose-averaging estimate in the example), a likelihood at a value below the point estimate, and a likelihood at a value above the point estimate. Specifically, the low value is taken to be the point estimate divided by 2.5 and the high value is taken to be the point estimate multiplied by 2.5.

(vii) The likelihood values are compared by calculating ratios of likelihoods, and then determining whether these likelihood-ratios (LR) exceed a critical value. Testing stops when the ratio of the likelihood for the point estimate exceeds each of the other likelihoods by a factor of 2.5, which is taken to indicate relatively strong statistical support for the point estimate. Therefore two likelihood-ratios (LRs) are calculated, a ratio of likelihoods for the point estimate and the point estimate divided by 2.5, and a ratio for the point estimate and the estimate times 2.5.

(viii) The calculations are easily performed in any spreadsheet with normal probability functions. The calculations are illustrated in Tables 6 and 7 in paragraph (m)(2)(iv) of this guideline, which is structured to promote spreadsheet implementation. The computation steps are illustrated using an example where the upper limit dose is 5000 mg/kg.

(A) Hypothetical example (Tables 6 and 7 in paragraph (m)(2)(iv) of this guideline). In the hypothetical example utilizing an upper boundary dose of 5000 mg/kg, the LR stopping criterion was met after nine animals had been tested. The first ''reversal'' occurred with the 3rd animal tested. The LR stopping criterion is checked when four animals have been tested following the reversal. In this example, the fourth animal tested following the reversal is the seventh animal actually tested. Therefore, for this example, the spreadsheet calculations are only needed after the seventh animal had been tested and the data could be entered at that time. Subsequently, the LR stopping criterion would have been checked after testing the seventh animal, the eighth animal, and the ninth. The LR stopping criterion is first satisfied after the ninth animal is tested in this example.

(*1*) Enter the dose-response information animal by animal.

(*i*) Column 1. Steps are numbered 1–15. No more than 15 animals may be tested.

(*ii*) Column 2. Place an I in this column as each animal is tested.

(*iii*) Column 3. Enter the dose received by the i^{th} animal.

(*iv*) Column 4. Indicate whether the animal responded (shown by an X) or did not respond (shown by an O).

(*2*) The nominal and actual sample sizes. The nominal sample consists of the two animals that represent the first reversal (here the second and third animals), plus all animals tested subsequently. Here, Column 5 indicates whether or not a given animal is included in the nominal sample.

(*i*) The nominal sample size (nominal *n*) appears in Row 16. This is the number of animals in the nominal sample. In the example, nominal *n* is 8.

(*ii*) The actual number tested appears in Row 17.

(*3*) Rough estimate of the LD_{50}. The geometric mean of doses for the animals in the current nominal sample is used as a rough estimate of the LD_{50} from which to gauge progress. In the table, this is called the ''dose-averaging estimator.'' It is updated with each animal tested. This average is restricted to the nominal sample in order to allow for a poor choice of initial test dose, which could generate either an initial string of responses or an initial string of nonresponses. (However, the results for all animals are used in the likelihood calculations for final LD_{50} calculation below.) Recall that the geometric mean of *n* numbers is the product of the *n* numbers, raised to a power of $1/n$.

(*i*) The dose-averaging estimate appears in Row 18 (e.g., $(175 * 550 * \ldots * 1750)^{1/8} = 1292.78$).

(*ii*) Row 19 shows the logarithm (base 10) of the value in Row 18 (e.g., $\log_{10} 1292.8 = 3.112$).

(*4*) Likelihood for the rough LD_{50} estimate.

(*i*) ''Likelihood'' is a statistical measure of how strongly the data support an estimate of the LD_{50} or other parameter. Ratios of likelihood values can be used to compare how well the data support different estimates of the LD_{50}.

(*ii*) In Column 8 calculate the likelihood for Step C's rough LD_{50} estimate. The likelihood (Row 21) is the product of likelihood contributions for individual animals (see paragraph (k)(2) of this guideline). The likelihood contribution for the i^{th} animal is denoted L_i.

(*iii*) Column 7. Enter the estimate of the probability of response at dose d_i, denoted P_i. P_i is calculated from a dose-response curve. Note that the parameters of a probit dose-response curve are the slope and the LD_{50}, so values are needed for each of those parameters. For the LD_{50} the dose-averaging estimate from Row 18 is used. For the slope in this example the default value of 2 is used. The following steps may be used to calculate the response probability P_i.

1. Calculate the base-10 log of dose d_i (Column 6).

2. For each animal calculate the z-score, denoted Z_i (not shown in the table), using the formulae

$sigma = 1 / slope$,

$Z_i = (\log_{10}(d_i) - \log_{10}(LD_{50})) / sigma$

For example, for the first animal (Row 1),

$sigma = 1 / 2$

$Z_1 = (2.243 - 3.112) / 0.500 = -1.738$

3. For the i^{th} dose the estimated response probability is

$P_i = F(Z_i)$

where F denotes the cumulative distribution function for the standard normal distribution (i.e., the normal distribution with mean 0 and variance 1).

For example (Row 1),

$P_1 = F(-1.738) = 0.0412$

The function F (or something very close) is ordinarily what is given for the normal distribution in statistical tables, but the function is also widely available as a spreadsheet function. It is available under different names, for example the @NORMAL function of Lotus 1-2-3 (see paragraph (n)(19) of this guideline) and the @NORMDIST function in Excel (see paragraph (n)(20) of this guideline). To confirm that you have used correctly the function available in your software, you may wish to verify familiar values such as $F(1.96) \approx 0.975$ or $F(1.64) \approx 0.95$.

(*iv*) Column 8. Calculate the natural log of the likelihood contribution ($\ln(L_i)$). L_i is simply the probability of the response that actually was observed for the i^{th} animal:

Responding animals: $\ln(L_i) = \ln(P_i)$

Non-responding animals: $\ln(L_i) = \ln(1 - P_i)$

Note that here the natural logarithm (ln) is used, whereas elsewhere the base-10 (common) logarithm was used. These choices are what are ordinarily expected in a given context.

The steps above are performed for each animal. Finally:

Row 20: Sum the log-likelihood contributions in Column 8.

Row 21: Calculate the likelihood by applying the exp function applied to the log-likelihood value in Row 20 (e.g., exp(-3.389) = $e^{-3.389}$ = 0.0337).

(*5*) Calculate likelihoods for two dose values above and below the rough estimate. If the data permit a precise estimate, then one expects the likelihood should be high if the estimate is a reasonable estimate of the LD_{50}, relative to likelihoods for values distant from this estimate. Compare the likelihood for the dose-averaging estimate (1292.8, Row 18) to values differing by a factor of 2.5 from that value (i.e., to 1292.8*2.5 and 1292.8/2.5). The calculations (displayed in Columns 9–12) are carried out in a fashion similar to those described above, except that the values 517.1 (=1292.8/2.5) and 3232.0 (=1292.8*2.5) have been used for the LD_{50}, instead of 1292.8. The likelihoods and log-likelihoods are displayed in Rows 20–21.

(*6*) Calculate likelihood-ratios. The three likelihood values (Row 21) are used to calculate two likelihood-ratios (Row 22). A likelihood-ratio is used to compare the statistical support for the estimate of 1292.8 to the support for each of the other values, 517.1 and 3232.0. The two likelihood-ratios are therefore:

LR1 = [likelihood of 1292.8] / [likelihood of 517.1]

= 0.0337 / 0.0080

= 4.21

and

LR2 = [likelihood of 1292.8] / [likelihood of 3232.0]

= 0.0337 / 0.0098

= 3.44

(*7*) Determine if the likelihood-ratios exceed the critical value. High likelihood-ratios are taken to indicate relatively high support for the point estimate of the LD_{50}. Both of the likelihood-ratios calculated in paragraph (m)(2)(viii)(A)(*6*) of this guideline (4.21 and 3.44) exceed the critical likelihood-ratio, which is 2.5. Therefore the LR stopping criterion is satisfied and testing stops. This is indicated by a TRUE in Row 24 and a note at the top of the example spreadsheet that the LR criterion is met. Determination of the point estimate and CI is carried out separately.

(B) [Reserved]

(3) **Performance of the UDP.** This section addresses choice of dose progression and initial dose level for the UDP and describes the performance of the test under a variety of circumstances. A companion document titled ''Toxicology Summary: Performance of the Up-and-Down Procedure'' provides assistance to the user in interpretation of the test results and is available on the ICCVAM web site at http://iccvam.niehs.nih.gov/methods/udpdocs/udprpt/udp_ciprop.htm. The statistical methods applied will depend upon the case into which the test response patterns fall (see Table 8 in paragraph (m)(3)(iii) of this guideline.

(i) Adjusting the dose progression and initial dose. For optimum performance of the UDP, the dose progression used should be based on an accurate prior estimate of *sigma*. The following two cases describe the outcome when an accurate estimate of *sigma* is not available. In addition, to account conservatively for any bias in the LD_{50} estimate, it is essential that dosing be initiated below the actual LD_{50}.

(A) Assumed *sigma* << true *sigma*: When the assumed *sigma* (i.e., the *sigma* on which the dose progression is based) is much smaller than the true *sigma* of the actual test population, the estimated LD_{50} may be ''biased'' in the direction of starting dose. For example, if the starting dose is less than the true LD_{50} of the test population, the estimated LD_{50} will generally be below the true LD_{50}. Also, if the starting dose is greater than the true LD_{50} of the test population, the estimated LD_{50} will tend to be greater than the true LD_{50}. To minimize the chance of overestimating the LD_{50} due to this bias, the UDP guideline recommends a choice of starting dose just below the assumed LD_{50}.

(B) Assumed *sigma* >> true *sigma*: If the assumed sigma on which the dose progression is based is much larger than the true *sigma* of the test population, the median estimated LD_{50} can be much larger or much smaller than the true LD_{50} depending on the starting dose. In this case, the LD_{50} can be estimated only within a range. (This is Case 3 described below.)

(ii) CI. Coverage of the CI is the probability that a calculated CI encloses the true LD_{50} for an experimental sample. Because the profile likelihood method is approximate, coverage of the CI does not always correspond to its nominal value. For example, coverage falls below 95% for populations with shallow slopes and is better than 95% for populations with steep slopes. In addition, the width of the CI is limited by the dose progression chosen. Generally, no type of CI would be more narrow than the dose progression.

(iii) Response Patterns. Data gathered under the UDP fall into one of five animal response patterns. The five types of animal response patterns, referred to as Case 1 through Case 5 in the following Table 8, can

be distinguished for the purpose of describing the performance of the UDP. These cases can be distinguished by looking at the experimental outcome (survival or death) as reflected in the AOT425StatPgm Data Grid or Report windows (see paragraph (n)(18) of this guideline). In considering these cases, note that doses can be repeated more than once in the course of sequential dosing.

Table 8.—Outcomes of the UDP: Cases and Confidence Intervals

Case #	Definition of Case	Approach Proposed	Possible Findings
1	No positive dose-response association. (1a) All animals tested in the study responded, or (1b) none responded, or (1c) the geometric mean dose is lower for animals that responded than for animals that did not respond.	LD_{50} cannot be calculated. CI not applicable.	Possible inferences: (1a) LD_{50} < lowest dose; (1b) LD_{50} > highest dose; (1c) reverse dose-response curve; unlikely test outcome. In case 1b, the highest dose tested is equivalent to a limit dose.
2	Multiple partial responses. One or more animals responded at a dose below some other dose where one or more did not respond. The conditions defining Case 1 do not hold. (The definition of Case 2 holds if there are 2 doses with partial responses, but holds in some other cases as well.)	Maximum likelihood estimate and profile likelihood computations of CI are straightforward.	The LD_{50} can be estimated and its CI calculated.
3	No intermediate response fractions. One or more test doses is associated with 0% response and one or more is associated with 100% response (all of the latter being greater than all of the former), and no test doses are associated with a partial response.	Lower bound = highest test dose with 0% response. Upper bound = lowest test dose with 100% response.	High confidence that the true LD_{50} falls between the two bounding doses. Any value of LD_{50} between highest dose with 0% response and lowest dose with 100% response is equally plausible.
4	One partial response fraction, first subcase. An intermediate partial response is observed at a single test dose. That dose is greater than doses associated with 0% response and lower than doses associated with 100% response.	The LD_{50} is set at the single dose showing partial response and its CI is calculated using profile likelihood method.	The LD_{50} can be estimated and its CI calculated.
5	One partial response fraction, second subcase. There is a single dose associated with partial response, which is either the highest test dose (with no responses at all other test doses) or the lowest test dose (with 100% response at all other test doses).	The LD_{50} is set at the dose with the partial response. A profile likelihood CI is calculated and may be finite or infinite.	The true LD_{50} could be at the boundary of the testing range with more or less confidence.

(n) References. The following references should be consulted for additional background material on this test guideline.

(1) Organization for Economic Cooperation and Development. OECD Guidelines for the Testing of Chemicals. Guideline 425: Acute Oral Toxicity—Up-and-Down Procedure. Adopted: December 2001.

(2) Organization for Economic Cooperation and Development. OECD Guidelines for the Testing of Chemicals. Guideline 420: Acute Oral Toxicity—Fixed Dose Method. Adopted: December 2001.

(3) Organization for Economic Cooperation and Development. OECD Guidelines for the Testing of Chemicals. Guideline 423: Acute Oral Toxcity—Acute Toxic Class Method. Adopted: December 2001.

(4) Dixon, W.J. and A.M. Mood. (1948). A Method for Obtaining and Analyzing Sensitivity Data. J. Amer. Statist. Assoc., 43, 109–126.

(5) Dixon, W.J. (1965). The Up-and-Down Method for Small Samples. J. Amer. Statist. Assoc. 60, 967–978.

(6) Dixon, W.J. (1991). Staircase Bioassay: The Up-and-Down Method. Neurosci. Biobehav. Rev., 15, 47–50.

(7) Dixon, W.J. (1991). Design and Analysis of Quantal Dose-Response Experiments (with Emphasis on Staircase Designs). Dixon Statistical Associates, Los Angeles CA, USA.

(8) Bruce, R.D. (1985). An Up-and-Down Procedure for Acute Toxicity Testing. Fundam. Appl. Tox., 5, 151–157.

(9) ASTM (1987). E 1163–87, Standard Test Method for Estimating Acute Oral Toxicity in Rats. American Society for Testing and Materials, Philadelphia Pa, USA.

(10) Lipnick, R.L., J.A. Cotruvo, R.N. Hill, R.D. Bruce, K.A. Stitzel, A.P. Walker, I. Chu, M. Goddard, L. Segal, J.A. Springer, and R.C. Myers. (1995). Comparison of the Up-and-Down, Conventional LD_{50} and Fixed Dose Acute Toxicity Procedures. Fd. Chem. Toxicol., 33, 223–231.

(11) Organization for Economic Cooperation and Development. (2000). Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation. Environmental Health and Safety Monograph Series on Testing and Assessment No. 19.

(12) Report of the International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity. NIH Publication No. 01-4499. Research Triangle Park, NC: National Institute of Environmental Health Sciences, August 2001.

(13) Guidance Document on Using *In Vitro* Data to Estimate *In Vivo* Starting Doses for Acute Toxicity. NIH Publication No. 01-4500. Research Triangle Park, NC: National Institute of Environmental Health Sciences, August 2001.

(14) Spielmann, H.E., M.Genschow, M. Leibsch, and W. Halle. (1999). Determination of the Starting Dose for Acute Oral Toxicity (LD_{50}) Testing in the Up-and-Down Procedure (UDP) from Cytotoxicity Data ATLA 27: 957–966.

(15) Chan, P.K. and A.W. Hayes. (1994). Chap. 16. Acute Toxicity and Eye Irritancy. *Principles and Methods of Toxicology*. Third Edition. A.W. Hayes, Editor. Raven Press, Ltd., New York, USA.

(16) Westat. (2001). Acute Oral Toxicity Software Program; AOT 425StatPgm; AOT425StatPgm Program User's Manual; and Simulation Results for the AOT425StatPgm Program. Reports prepared for U.S.

E.P.A. under Contract 68–W7–0025, Task Order 5-03. Currently available at web site: http://iccvam.niehs.nih.gov/methods/udpdocs/udprpt/udp_ciprop.htm

(17) Rosenberger, W.F., Flournoy, N. and Durham, S.D. (1997). Asymptotic Normality of Maximum Likelihood Estimators from Multiparameter Response-driven Designs. Journal of Statistical Planning and Inference 60, 69–76.

(18) Jennison, C. and B.W. Turnbull. (2000). *Group Sequential Methods with Applications to Clinical Trials*. Chapman & Hall/CRC: Boca Raton, FL.

(19) Lotus Development Corporation. (1999). Lotus® 1-2-3. Version 9.5, Millennium edition. Cambridge, MA, USA.

(20) Microsoft Corporation. (1985–1997). Microsoft® Excel. Version 5.0 or later. Seattle, WA, USA.

[This Page Intentionally Left Blank]

APPENDIX B6

OECD GUIDANCE DOCUMENT 24: ACUTE ORAL TOXICITY TESTING

[This Page Intentionally Left Blank]

Unclassified **ENV/JM/MONO(2001)4**

Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development **23-Jul-2001**

English - Or. English

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

ENV/JM/MONO(2001)4
Unclassified

OECD SERIES ON TESTING AND ASSESSMENT
Number 24

GUIDANCE DOCUMENT ON ACUTE ORAL TOXICITY TESTING

English - Or. English

JT00111082

ENV/JM/MONO(2001)4

OECD Environment, Health and Safety Publications

Series on Testing and Assessment

N° 24

GUIDANCE DOCUMENT ON ACUTE ORAL TOXICITY TESTING

Environment Directorate

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Paris

June 2001

Also published in the Series on Testing and Assessment

No. 1, *Guidance Document for the Development of OECD Guidelines for Testing of Chemicals (1993; reformatted 1995)*

No. 2, *Detailed Review Paper on Biodegradability Testing (1995)*

No. 3, *Guidance Document for Aquatic Effects Assessment (1995)*

No. 4, *Report of the OECD Workshop on Environmental Hazard/Risk Assessment (1995)*

No. 5, *Report of the SETAC/OECD Workshop on Avian Toxicity Testing (1996)*

No. 6, *Report of the Final Ring-test of the Daphnia magna Reproduction Test (1997)*

No. 7, *Guidance Document on Direct Phototransformation of Chemicals in Water (1997)*

No. 8, *Report of the OECD Workshop on Sharing Information about New Industrial Chemicals Assessment (1997)*

No. 9, *Guidance Document for the Conduct of Studies of Occupational Exposure to Pesticides During Agricultural Application (1997)*

No. 10, *Report of the OECD Workshop on Statistical Analysis of Aquatic Toxicity Data (1998)*

No. 11, *Detailed Review Paper on Aquatic Testing Methods for Pesticides and industrial Chemicals (1998)*

No. 12, *Detailed Review Document on Classification Systems for Germ Cell Mutagenicity in OECD Member Countries (1998)*

No. 13, *Detailed Review Document on Classification Systems for Sensitising Substances in OECD Member Countries (1998)*

No. 14, *Detailed Review Document on Classification Systems for Eye Irritation/Corrosion in OECD Member Countries (1998)*

No. 15, *Detailed Review Document on Classification Systems for Reproductive Toxicity in OECD Member Countries (1998)*

No. 16, *Detailed Review Document on Classification Systems for Skin Irritation/Corrosion in OECD Member Countries (1998)*

No. 17, *Environmental Exposure Assessment Strategies for Existing Industrial Chemicals in OECD Member Countries (1999)*

No. 18, *Report of the OECD Workshop on Improving the Use of Monitoring Data in the Exposure Assessment of Industrial Chemicals (2000)*

No. 19, *Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation (2000)*

No. 20, *Revised Draft Guidance Document for Neurotoxcity Testing (2000)*

No. 21, *Detailed Review Paper: Appriasal of Test Methods for Sex Disrupting Chemicals (2000)*

No. .22, *Guidance Document for the Performance of Out-door Monolith Lysimeter Studies (2000)*

No. 23, *Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures (2000)*

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 29 industrialised countries in North America, Europe and the Pacific, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised Committees and subsidiary groups composed of Member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's Workshops and other meetings. Committees and subsidiary groups are served by the OECD Secretariat, located in Paris, France, which is organised into Directorates and Divisions.

The work of the OECD related to chemical safety is carried out in the **Environment, Health and Safety Programme**. As part of its work on chemical testing, the OECD has issued several Council Decisions and Recommendations (the former legally binding on Member countries), as well as numerous Guidance Documents and technical reports. The best known of these publications, the **OECD Test Guidelines**, is a collection of methods used to assess the hazards of chemicals and of chemical preparations such as pesticides. These methods cover tests for physical and chemical properties, effects on human health and wildlife, and accumulation and degradation in the environment. The OECD Test Guidelines are recognised world-wide as the standard reference tool for chemical testing.

More information about the Environment, Health and Safety Programme and its publications (including the Test Guidelines) is available on the OECD's World Wide Web site (see next page).

The Environment, Health and Safety Programme co-operates closely with other international organisations. This document was produced within the framework of the Inter-Organisation Programme for the Sound Management of Chemicals (IOMC).

The Inter-Organization Programme for the Sound Management of Chemicals (IOMC) was established in 1995 by UNEP, ILO, FAO, WHO, UNIDO and the OECD (the Participating Organisations), following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. UNITAR joined the IOMC in 1997 to become the seventh Participating Organisation. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

INTRODUCTION

1. OECD Guidelines for the Testing of Chemicals are periodically reviewed in the light of scientific progress or changing assessment practices. The conventional acute oral toxicity test (formerly OECD Test Guideline 401) is the most heavily criticised test in terms of animal welfare and this concern was the driving force behind the development of three alternative tests for acute oral toxicity (Test Guideline 420, 423, 425). Anticipating the presence of validated alternatives, Member countries took the initiative to plan the deletion of Guideline 401.

2. A Nominated Expert Meeting (Rome 1998) and an Expert Consultation Meeting, (Arlington 1999) were convened to determine the acute oral toxicity data requirement needs of Member countries and to assess the capabilities of the alternatives to meet these needs. On the basis of these technical discussions, the 29th Joint Meeting concluded in June 1999 that not all data needs could be met by the alternatives (and not always by Guideline 401). The Joint Meeting decided that Guidelines 420, 423 and 425 should be revised to meet regulatory needs of the Member countries including, where possible, the provision of confidence intervals and the slope of the dose response curve, to support classification and assessment of acute toxicity at 5 and at 5000 mg/kg, and should include the use of a single sex, appropriate statistical methods and, to the extent feasible, a reduction in the number of animals used and the introduction of refinements to reduce the pain and distress of the animals. The guidelines should also be able to allow the classification of substances according to the Globally Harmonised System (GHS) for the classification of chemicals which cause acute toxicity (1).

3. The revision of Guidelines 420, 423 and 425 was completed in 2000 following a second Expert Consultation Meeting (Paris, 2000) and the process of deletion of guideline 401 was started.

PURPOSE

4. The purpose of this Guidance Document is to provide information for both the regulated community and regulators to assist with the choice of the most appropriate Guideline to enable particular data requirements to be met while reducing the number of animals used and animal suffering. The Guidance Document also contains additional information on the conduct and interpretation of Guidelines 420, 423 and 425.

DATA NEEDS

5. Acute oral toxicity data are used to satisfy hazard classification and labelling requirements, for risk assessment for human health and the environment, and when estimating the toxicity of mixtures. The provision of either a point estimate of the LD_{50} value or range estimate of the LD_{50} generally meets the acute oral toxicity data requirements for classification for all regulatory authorities in the areas of industrial chemicals, consumer products and for many pesticide applications. OECD document "Revised Analysis of Responses Received from Member Countries to the Questionnaire on Data Requirements for Acute Oral Toxicity" provides an overview of acute toxicity data requirements applicable in 1999 (2). The data needs of the majority of Member countries can also be met with the imposition of a limit dose of 2000 mg/kg. However, several countries have a requirement for information on toxicity at dose levels in the range 2000 to 5000 mg/kg for substances with LD_{50} values in excess of 2000 mg/kg. Although many authorities find it acceptable to use data from observations made at doses of 2000 mg/kg or below, as

described in the GHS classification criteria (which includes a 2000-5000 mg/kg category), testing in this range may be necessary to meet the needs of a few regulatory authorities. For example, some authorities regulating consumer products and pesticides need a point estimate of LD_{50} and confidence intervals, and information on toxicity at levels up to or above 5000 mg/kg. These authorities use LD_{50} data in this way for assessment of risk to humans and also for risk assessments for environmental effects to avoid the need for further animal studies on pesticide products. Furthermore, at least one country has a need for a test at 5000 mg/kg for biological and safer pesticides and products to which the general public are exposed, to provide characterisation of acute toxicity and to support bridging across data sets for structurally related substances, again to eliminate or minimise the requirements for additional animal testing. For reasons of animal welfare concern, testing of animals in GHS category 5 ranges (2000-5000mg/kg) is discouraged and should only be considered when there is a strong likelihood that results of such a test have a direct relevance for protecting human or animal health or the environment.

6. Some national and international regulatory systems estimate the toxicity of mixtures from calculations using weighted averages of the LD_{50} point estimate of the components when actual data on the mixture are not available. The resulting calculated toxicity values are used for hazard classification of mixtures. A dose response curve is also sometimes needed for extrapolation and a reliable identification of hazard and risk posed by mixtures, to avoid testing each mixture and thus to allow a significant saving of animal use. At present, agreed approaches for estimating the toxicity of mixtures using range data are only accepted in the EU and in some other countries. However, the OECD Expert Group on Hazard Classification Criteria for Mixtures has recently agreed that mixtures can be classified using either point or range estimates of the LD50 of each component (3).

7. Acute oral toxicity testing by OECD methods is not required for pharmaceuticals. Pharmaceutical methods are specified by the International Committee on Harmonisation (ICH). In some specific cases such as imaging and antineoplastic agents, estimates of acute toxicity are needed to support single dose studies in man. These studies call for testing to fully characterise the toxicity in the low toxicity region and may involve doses above 2000 mg/kg. However, the study designs for these special purpose studies are different from any of the current OECD acute toxicity guidelines.

COMPARISON OF GUIDELINES 420, 423 AND 425

Outline Of The Methodology

8. All of the guidelines involve the administration of a single bolus dose of test substance to fasted healthy young adult rodents by oral gavage, observation for up to 14 days after dosing, recording of body weight and the necropsy of all animals. Doses may be administered based on a constant volume or a constant concentration depending upon the needs of the toxicologist and the regulatory authorities. Some authorities prefer that substances sold to the public should be tested as constant concentration unless the volumes are too small to administer accurately. Since the effects at the same dose may be different if the materials are diluted, it is important for the toxicologist to consider how the information will be used. If the material will primarily be used diluted in mixtures, then constant volume may be appropriate. On the other hand, if the material is to be used neat, particularly if it may be irritating, the use of constant concentration will be more appropriate (4)(5).

9. Each animal should be selected from the available animals in a random fashion on the day of dosing. In recognition of the fact that most animal suppliers do not indicate littermates, the guidelines do not call for randomizing animals from a single litter across dose groups. Females should be nulliparous

and non-pregnant. At the commencement of its dosing, each animal should be between 8 and 12 weeks old and its weight should fall in an interval within ±20% of the mean weight of all previously dosed animals taken on their day of dosing. As the mean weight will increase as the animals age, this method tends to correct for the change in animals weights with time. In order to conform to these age and weight requirements at the start of dosing of each animal, it may be necessary to order animals sequentially as the tests can sometimes take several weeks to complete. The primary endpoint for Guidelines 423 and 425 is mortality, but for Guideline 420 it is the observation of clear signs of toxicity (termed: evident toxicity).

10. **Guideline 420:** A sighting study is included for Guideline 420 in order to choose an appropriate starting dose and to minimise the number of animals used. Pre-specified fixed doses of 5, 50, 300 or 2000 mg/kg are used both in the sighting study and the main study. There is an option to use an additional dose level of 5000 mg/kg, but only when justified by a specific regulatory need. Groups of animals are dosed in a stepwise procedure, with the initial dose being selected as the dose expected to produce some signs of toxicity. Further groups of animals may be dosed at higher or lower fixed doses, depending on the presence of signs of toxicity, until the study objective is achieved; that is, the classification of the test substance based on the identification of the dose(s) causing evident toxicity, except when there are no effects at the highest fixed dose.

11. **Guideline 423:** Pre-specified fixed doses of 5, 50, 300 or 2000 mg/kg are used. There is an option to use an additional dose level of 5000 mg/kg, but only when justified by a specific regulatory need. Groups of animals are dosed in a stepwise procedure, with the initial dose being selected as the dose expected to produce mortality in some animals. Further groups of animals may be dosed at higher or lower fixed doses, depending on the presence of mortality, until the study objective is achieved; that is, the classification of the test substance based on the identification of the dose(s) causing mortality, except when there are no effects at the highest fixed dose.

12. **Guideline 425:** This is also a stepwise procedure, but uses single animals, with the first animal receiving a dose just below the best estimate of the LD_{50}. Depending on the outcome for the previous animal, the dose for the next is increased or decreased, usually by a factor of 3.2. This sequence continues until there is a reversal of the initial outcome (i.e., the point where an increasing dose results in death rather than survival, or decreasing dose results in survival rather than death); then, additional animals are dosed following the up-down principle until a stopping criterion is met. If there is no reversal before reaching the selected upper (2000 or 5000 mg/kg) limit dose, then no more than a specified number of animals are dosed at the limit dose. The option to use an upper limit dose of 5000 mg/kg should be taken only when justified by a specific regulatory need.

Animal Welfare Considerations

13. All three Guidelines provide significant improvements in the number of animals used in comparison to Guideline 401, which required 20 animals in a test at least. In addition, they all contain a requirement to follow the OECD Guidance Document on Humane Endpoints (6) which should reduce the overall suffering of animals used in this type of toxicity test. Furthermore, Guideline 420 has as its endpoint evident toxicity rather than mortality and uses a sighting study to minimize the numbers of animals and Guideline 425 has a stopping rule which limits the number of animals in a test.

14. **Guideline 420:** Groups of five young adult animals of one sex are dosed per step in the main study. Single animals are used per step in the sighting study. Regulatory experience and statistical modelling has shown that most tests are likely to be completed with either one or two sighting study steps and one main study step, thus using between 5 and 7 animals. Up to 5 animals are used in a limit test.

15. **Guideline 423:** This test uses groups of 3 animals of one sex per step. Regulatory use of this Guideline demonstrates that the average number of animals used is 7. Up to 6 animals are used in a limit test.

16. **Guideline 425:** This test uses single animals of one sex. Statistical modelling indicates that the average number of animals used in this test is about 6-9. Up to 5 animals are used in a limit test.

17. The following estimates of the number of treatment related deaths for tests conducted on substances with LD_{50} values below 5000 mg/kg are based on practical experience and validation studies using earlier versions of these guidelines and statistical modelling.

- **Guideline 420:** typically 1 animal can be expected to die on test.
- **Guideline 423:** 2-3 animals per test can be expected to die in a full test.
- **Guideline 425:** the expected number of deaths is between 2 and 3.

18. For all three guidelines, careful clinical observations should be made at least twice on the day of dosing or more frequently when indicated by the response of the animals to the treatment, and at least once daily thereafter. Additional observations are made if the animals continue to display signs of toxicity. Observations include changes in skin and fur, eyes and mucous membranes, and also respiratory, circulatory, autonomic and central nervous systems, and somatomotor activity and behaviour pattern. Guidance on clinical signs can be found in Chan and Hayes (5). Animals that are moribund or suffering severe pain and distress must be humanely killed. Guidance on clinical signs and objective measurements that are indicative of impending death and/or severe pain and/or distress is available in an OECD Guidance Document (6). Humanely killed animals are considered in the interpretation of the results in the same way as animals that died on test.

Information Provided By Each Method

19. Test Guidelines 420 and 423 provide a range estimate of the LD_{50}; the ranges are defined by cut-off values of the applied classification system and not as a calculated lower and upper level. In the case of Test Guideline 420 this range is inferred from the fixed dose which produces evident toxicity. Guideline 425 provides a point-estimate of the LD_{50} value with confidence intervals.

20. The results of tests conducted according to Guideline 425 will allow a test substance to be classified according to all the systems in current use, including the new GHS. Test Guidelines 420 and 423 have now been revised to allow classification according to the new GHS. However, in order to cover the transition period until the global implementation of the GHS both Guidelines also allow classification according to existing systems as shown in Annex 1 and 2.

Limitations Of The Methods

21. Validations against actual data and statistical simulations identified areas where all three methods may have outcomes which result in a more or less stringent classification than that based on the "true" LD_{50} value (as obtained by the deleted guideline 401). Comparative statistical analysis (see Annex 3) indicated that all are likely to perform poorly for chemicals with shallow dose-response slopes. For all methods, the study outcome is likely to be influenced by the choice of starting dose level(s), relative to the "true" LD_{50} value, especially in the case of shallow slopes. Because Guideline 420 uses evident toxicity as

an endpoint instead of death, information on toxic effects seen only at dose levels close to a lethal dose will not always be obtained (7).

22. Unusually test substances may cause delayed deaths (5 days or more after test substance administration). Substances which cause delayed deaths have an impact on the practicality of conducting a study to Guideline 425 where the duration of testing will be significantly longer compared with other test methods. However, both in Guideline 420 and 423, the finding of a delayed death may require additional lower dose levels to be used or a study to be repeated.

OPTIMISING THE PERFORMANCE OF THE TEST

23. Each guideline provides procedures to assist in selecting the starting dose, particularly in the event that minimal prior information on the substance itself is available. All available information on the test substance must be made available to the testing laboratory and should be considered prior to conducting the study. Such information will include, for example, the identity and chemical structure of the substance; its physico-chemical properties; the result of any other *in vivo* or *in vitro* toxicity tests on the substance; toxicological data on structurally related substances; the anticipated use(s) of the substance; and the likely regulatory data requirements. This information is necessary to satisfy all concerned that the test is relevant for the protection of human and animal health and mammalian wildlife, to select the most appropriate test to satisfy regulatory requirements and will help in the selection of the starting dose.

24. For all three methods the efficiency of the test, in terms of reliability and numbers of animals used, is optimised by the choice of a starting dose close to (423) or just below (425) the actual LD_{50} or the lowest dose producing evident toxicity (420). When this type of information is not available, all three Guidelines include advice on the starting dose level which should be used to minimise the possibility of biased outcome and adverse effects on animal welfare. As a general principle it is suggested that a starting dose is selected that is slightly lower than the best estimate of the LD_{50} based on available evidence.

25. The limit test is an efficient way to characterise substances of low toxicity when there is sufficient information available indicating that the toxic dose is higher that the limit dose. Each method provides a limit test suitable to the design of the main study. A Limit Test should be conducted only when there are strong indications that the test substance is of low or negligible acute toxicity.

USE OF A SINGLE SEX

26. Guidelines 420, 423 and 425 are conducted using a single sex in order to reduce variability and as a means of minimising the number of animals used. Normally females are used. This is because literature surveys of conventional LD_{50} tests show that usually there is little difference in sensitivity between the sexes but, in those cases where differences were observed, females were generally slightly more sensitive (8). Although the use of a single sex (females) also contributes to a further decrease in the use of animals in testing, theoretically this may lead to an oversupply of the other sex (males). However, currently the use of males in experimental animal tests clearly exceeds that of females and, thus, the preference for females in acute toxicity testing may well result in a better overall balance of the use of both genders. For chemicals which are direct acting in their toxic mechanism, this may be because female rats have a lower detoxification capacity than males, as measured by specific activity of phase I and II enzymes. However, all available information should be evaluated, for example on chemical analogues and the results of testing for other toxicological endpoints on the chemical itself, as this may indicate that

males may be more sensitive than females. Knowledge that metabolic activation is required for a chemical's toxicity can also indicate that males may be the more sensitive sex.

27. Occasionally, the results of subsequent testing, for example a sub-chronic test, may raise concerns that the more sensitive sex had not been used. In such cases, and only when considerable differences between the sexes are suspected, it may be necessary to conduct another full acute oral toxicity study in the second sex. This is preferable to conducting confirmatory testing in a small group of animals of the second sex as a late satellite to the original test because there is a strong possibility that this would produce results that are difficult to interpret. The impact of conducting a second full test on the overall number of animals used in acute toxicity testing should be small because re-testing is anticipated to be infrequent and the results of the test in one sex, together with data from any subsequent studies, will greatly assist in the selection of starting doses closer to the LD50 in the second test.

REFERENCES

(1) OECD (1998) A Harmonised Integrated Hazard Classification System for Human Health and Environmental Effects of Chemicals. ENV/MC/CHEM/HCL(98)15/REV2.

(2) OECD(1999). Revised Analysis of Responses Received from Member Countries to the Questionnaire on Data Requirements for Acute Oral Toxicity. Environmental Health and Safety Monograph Series on Testing and Assessment N° 26 [http:// www.oecd.org/ehs/test/monos/htm.]

(3) OECD(2000).Revised Step 2 Proposal for Harmonised Classification Criteria for Mixtures. ENV/JM/HCL(2000)9/REV8.

(4) The Conduct of Acute Toxicity Studies, September 1977, E.P.A. document # 737-R-97-002.

(5) Chan P.K. and Hayes A.W. (1994). Acute Toxicity and Eye Irritancy. In: Principles and Methods of Toxicology. A.W Hayes, Ed. Raven Press Ltd, New York.

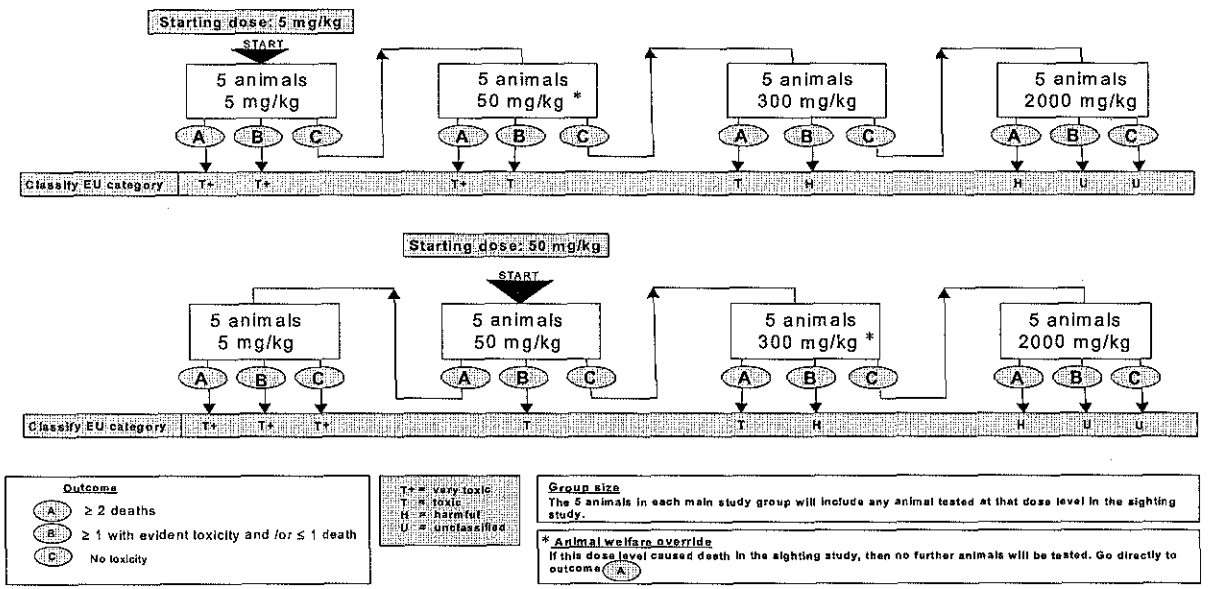
(6) OECD (2000). Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation. Environmental Health and Safety Monograph Series on Testing and Assessment No. 19, September 2000. [http://www.oecd.org/ehs/test/monos/htm]

(7) Van der Heuvel, M.J., Clark, D.G., Fielder, R.J., Koundakjian, P.P., Olivier, G.J.A., Pelling, D., Tomlinson, N.J. and Walker, A.P. (1990). The international validation of a fixed-dose procedure as an alternative to the classical LD_{50} test. Fd. Chem. Toxicol. 28, 469-482.

(8) Lipnick R.L, Cotruvo J.A., Hill R.N., Bruce R.D., Stitzel K.A., Walker A.P., Chu, I., Goddard M., Segal L., Springer J.A., Myers R.C. (1995). Comparison of the Up-and-Down, Conventional LD50 and Fixed Dose Acute Toxicity Procedures. Fd. Chem. Toxicol. 33, 223-231.

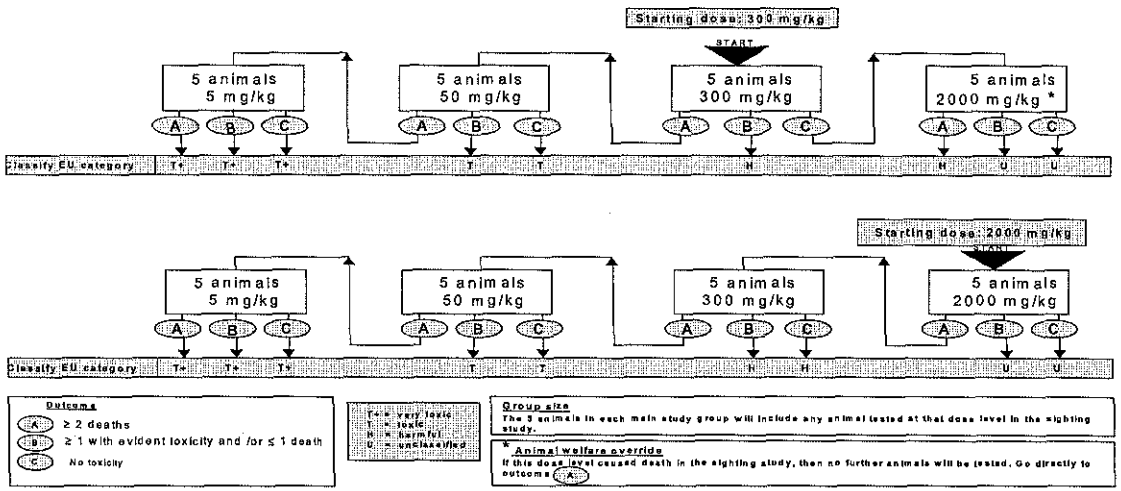
ANNEX 1

TEST GUIDELINE 420 MAIN STUDY: CLASSIFICATION ACCORDING TO THE CURRENTLY STILL APPLICABLE EU SCHEME TO COVER THE TRANSITION PERIOD UNTIL FULL IMPLEMENTATION OF THE GLOBALLY HARMONISED CLASSIFICATION SYSTEM (GHS)



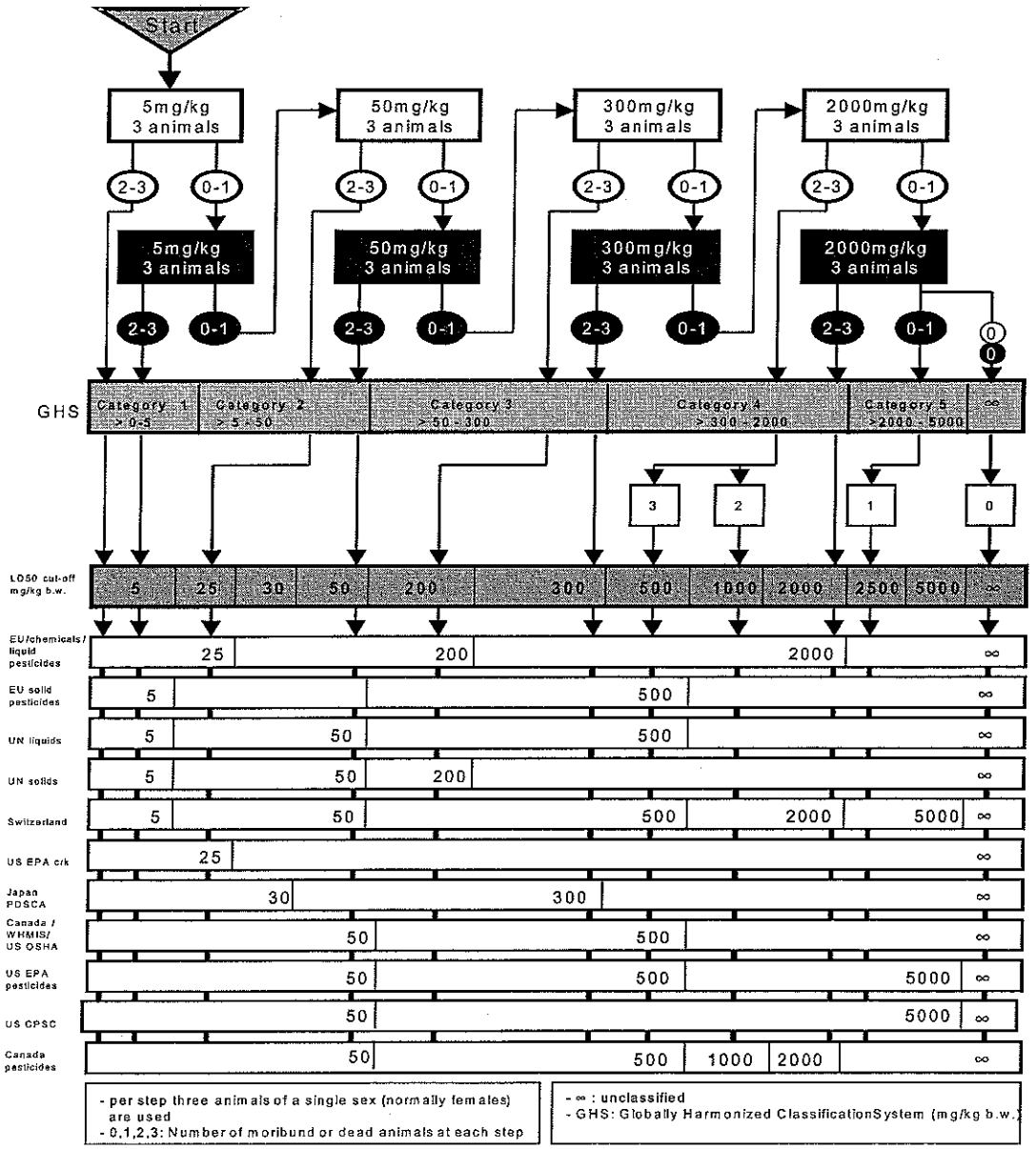
ANNEX 1 (continued)

TEST GUIDELINE 420 MAIN STUDY: CLASSIFICATION ACCORDING TO THE CURRENTLY STILL APPLICABLE EU SCHEME TO COVER THE TRANSITION PERIOD UNTIL FULL IMPLEMENTATION OF THE GLOBALLY HARMONISED CLASSIFICATION SYSTEM (GHS)



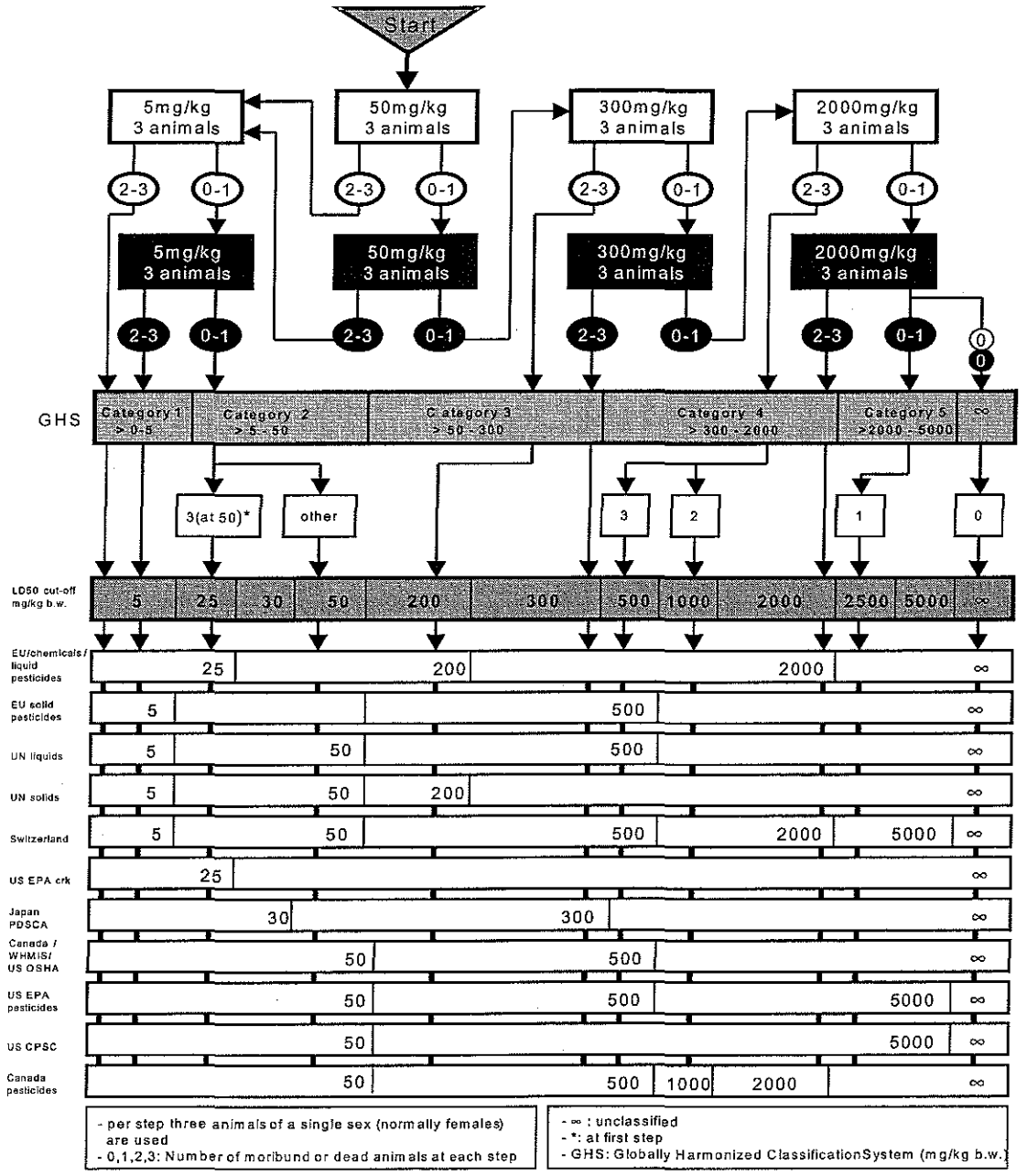
ANNEX 2

TEST GUIDELINE 423: CLASSIFICATION ACCORDING TO CURRENTLY STILL APPLICABLE CLASSIFICATION SCHEMES TO COVER THE TRANSITION PERIOD UNTIL FULL IMPLEMENTATION OF THE GLOBALLY HARMONISED CASSIFICATION SYSTEM (GHS)



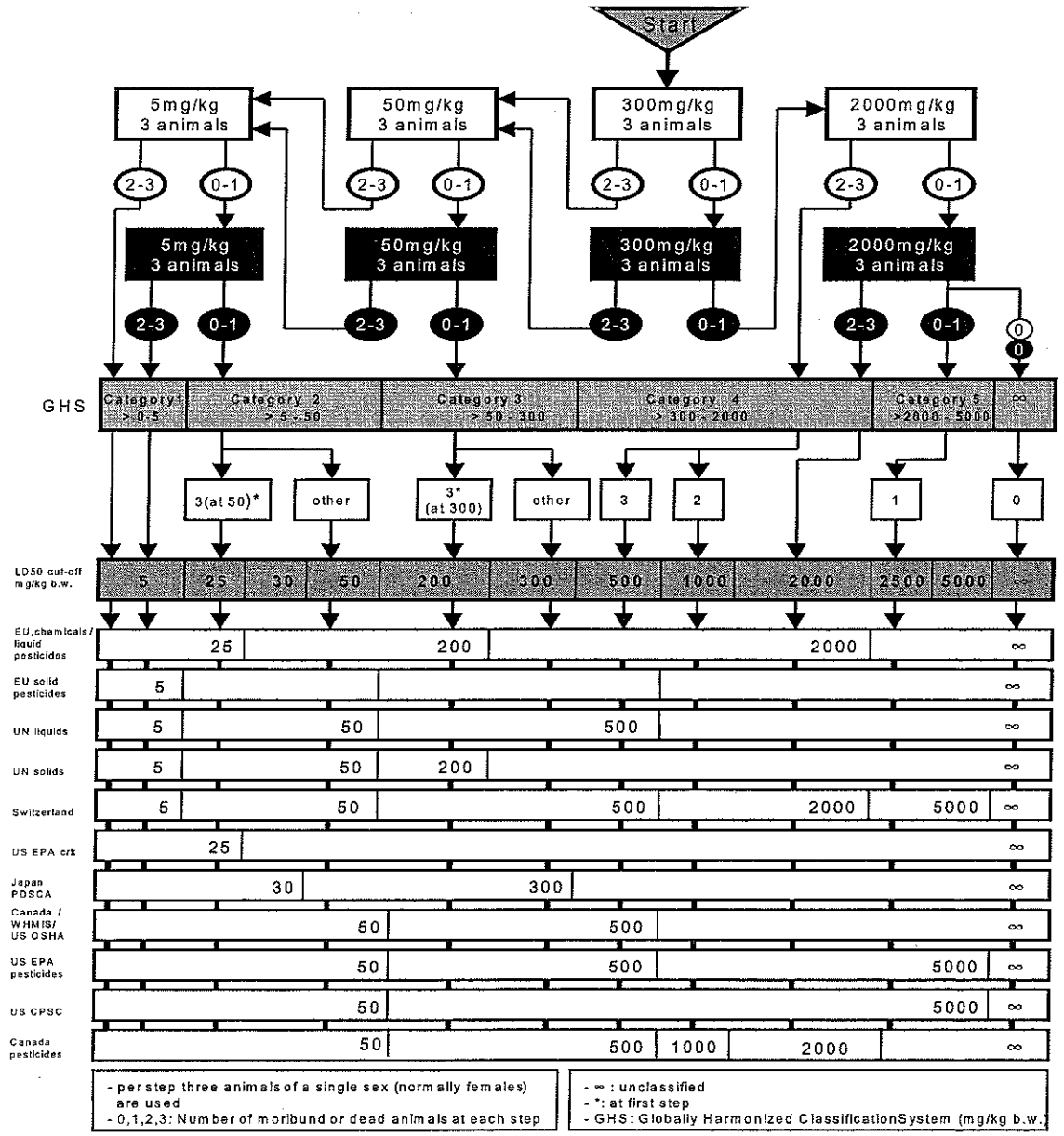
ANNEX 2 (continued 1)

TEST GUIDELINE 423: CLASSIFICATION ACCORDING TO CURRENTLY STILL APPLICABLE CLASSIFICATION SCHEMES TO COVER THE TRANSITION PERIOD UNTIL FULL IMPLEMENTATION OF THE GLOBALLY HARMONISED CASSIFICATION SYSTEM (GHS)



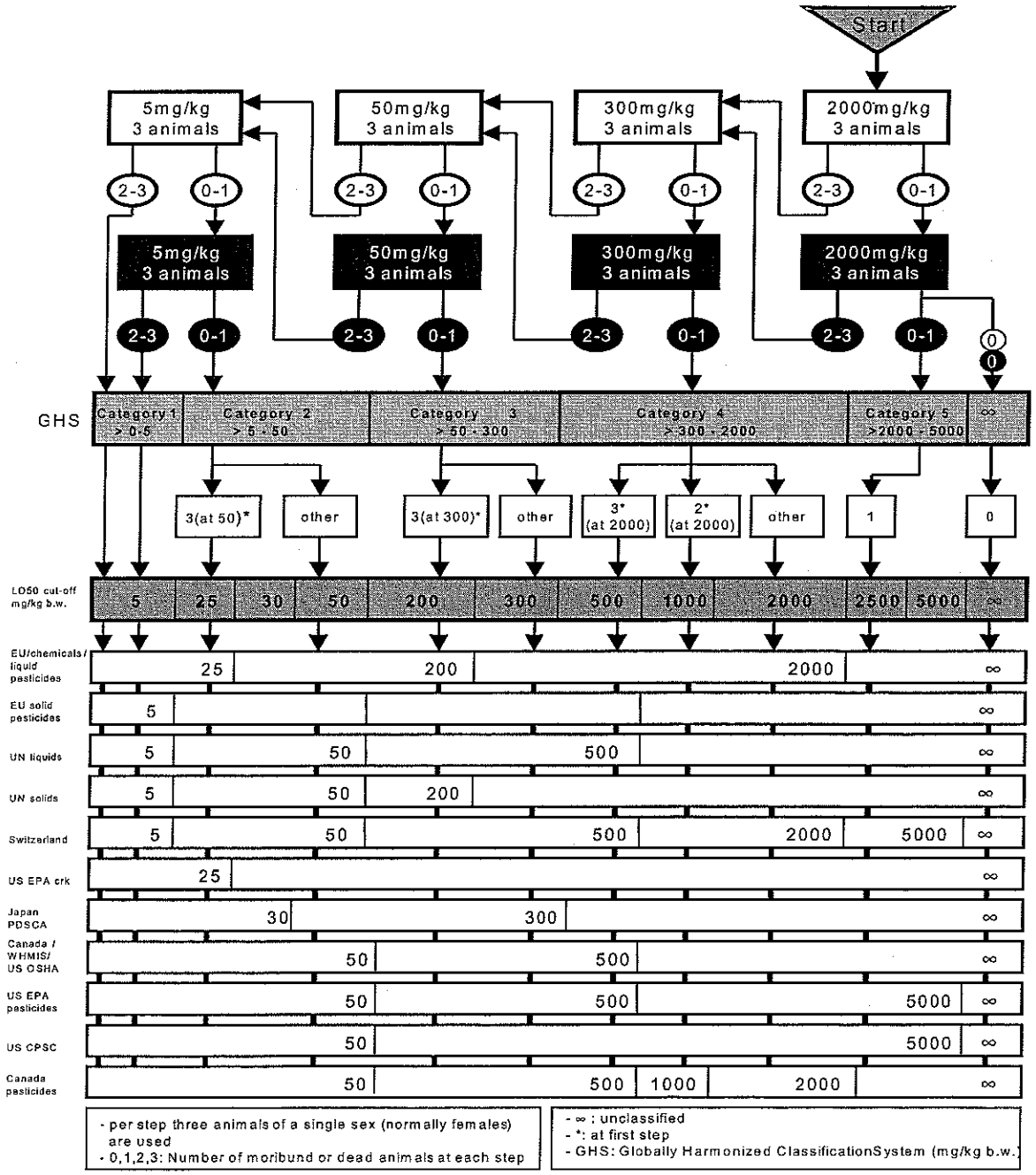
ANNEX 2 (continued 2)

TEST GUIDELINE 423: CLASSIFICATION ACCORDING TO CURRENTLY STILL APPLICABLE CLASSIFICATION SCHEMES TO COVER THE TRANSITION PERIOD UNTIL FULL MPLEMENTATION OF THE GLOBALLY HARMONISED CASSIFICATION SYSTEM (GHS)



ANNEX 2 (continued 3)

TEST GUIDELINE 423: CLASSIFICATION ACCORDING TO CURRENTLY STILL APPLICABLE CLASSIFICATION SCHEMES TO COVER THE TRANSITION PERIOD UNTIL FULL IMPLEMENTATION OF THE GLOBALLY HARMONISED CASSIFICATION SYSTEM (GHS)



ANNEX 3

STATISTICAL BASIS FOR ESTIMATING ACUTE ORAL TOXICITY COMPARISON OF OECD GUIDELINES 420, 423 AND 425

INTRODUCTION

1. This document describes the statistical strengths and limitations of the various methods for accurately determining a point estimate of the LD_{50}, confidence limits around the point estimate of LD_{50}, and information on the dose-effect response. In this context, a dose-response curve applies to the estimation of lethality and a dose-effect response applies to the estimation of the change in the variety and distribution of all other types of toxicological signs with the change in dose. By design not all of the guidelines will provide estimates for all of these endpoints. This document allows the reader to quickly identify the tests that will meet his or her particular needs.

2. The statistical basis for all test methods is that lethality is a quantal response. Its measurement will give rise to a frequency distribution of responses reflecting the composite tolerances of the test population upon exposure to graded doses of the test chemical. In practice, most chemicals give rise to an approximately lognormal distribution of deaths versus dose, skewed toward hypersensitivity. When this frequency population is transformed to a logarithmic abscissa, a (symmetric) normal distribution generally results that can be characterized by two parameters, the median and the standard deviation, SD. The median is the dose at which 50% of the animals are killed by the test chemical and is called the LD_{50}. Not all animals will react in the same way to the chemical and thus SD represents the square root of the variance of the test populations' response to the chemical. The dose-response curve is sigmoidal in nature and represents the cumulative response of the test animals to the chemical. The inflection point of this sigmoidal curve coincides with the LD_{50} for the test population.

3. What follows is a brief description of the mathematical and biological principles underlying each acute oral toxicity method followed by a listing of how each test estimates or does not estimate the specific parameters mentioned above.

GUIDELINE 420 :FIXED DOSE PROCEDURE

Principles Underlying The Test Method

4. The Fixed Dose Procedure (FDP) is a method for assessing acute oral toxicity that involves the identification of a dose level that causes evidence of non-lethal toxicity (termed *evident* toxicity) rather than a dose level that causes lethality. *Evident toxicity* is a general term describing clear signs of toxicity following administration of test substance, such that an increase to the next highest fixed dose would be expected to result in the development of severe toxic signs and probably mortality.

5. Underpinning the FDP is a belief that the toxic profile of a substance can be characterized with sufficient reliability for most regulatory situations without the need for the identification of a lethal dose. That is, observations made at non-lethal doses will allow substances to be ranked, or classified, according to their acute toxicity, provide information to aid dose level selection for repeat dose studies and provide hazard data for use in a risk assessment. The original FDP was subject to a number of validation and comparison studies, which showed that classification outcome was similar to that based on the outcome of traditional tests for determining an LD_{50} value (1)(2)(3)(4)(5).

6. Fixed dose levels of 5, 50, 300 and 2000 mg/kg and rules for the sequential procedure were adopted following a rigorous analysis using a statistical model (6)(7). The analysis predicted the classification outcome (according to the EU scheme and the lethality-based GHS), numbers of animals used and number of substance-related deaths using a number of FDP design options for substances with a range of LD_{50} values and dose response slopes for lethality. On the basis of this analysis, the design of the FDP was optimised with respect to classification performance and animal welfare.

7. The statistical modelling showed that the FDP produces classification outcomes similar to that based on the LD_{50} value for substances with a steep (greater than 2) dose response curve for mortality. For substances with a relatively shallow (less than 2) dose response curve there is an increasing probability the FDP will produce a more stringent classification than that based on the LD_{50} value; however, the risk of a less stringent classification than that based on the LD_{50} value is negligible. The influence of the choice of starting dose on the classification outcome, which can be a problem with sequential procedures, is negligible.

Point Estimate of LD_{50}

8. The FDP is not designed to determine a point estimate of LD_{50}. However, an approximate LD_{50} range can be inferred from the classification outcome. The ability of the FDP to correctly classify (i.e. assign to an LD_{50} range) is discussed above.

Confidence Limits on the Estimate of LD_{50}

9. The FDP is not designed to determine a point estimate of LD_{50}, or confidence limits on the estimate of the LD_{50}.

Dose-Effect Curve

10. Since lethality is not the preferred endpoint for the FDP, information on toxicological effects seen only at dose levels close to a lethal dose will not always be available. However, it has been shown in a number of validation and comparative studies (1)(2)(3)(4)(5)(6) that while there were instances where clinical signs observed in FDP tests differed from those observed in traditional LD_{50} tests, in only a few cases were these meaningful. In the majority of cases, the clinical signs not observed in the FDP tests were non-specific signs of approaching death.

GUIDELINE 423 : ACUTE TOXIC CATEGORY METHOD

Principles Underlying The Test Method

11. The acute toxic category (ATC) method allows for the allocation of chemical substances to all classification systems currently in use (e.g.. the LD_{50} is between 50 and 500 mg/kg body weight) (8)(9). It is a group sequential procedure using three animals of one sex per step. Four pre-identified starting doses are possible.

12. The ATC Method is based on the probit model; i.e., the dose-response relationship follows the Gaussian distribution for log-dose values with two parameters, the mean (LD_{50}) and the slope in probit units based on the log-scaled dose-axis (logarithm according to base 10). Then, following the test scheme of the method, expected probabilities of a correct, of a lower and of a more stringent classification in dependence on the true oral LD_{50} value of a substance and its slope can be derived.

13. The test doses were selected with respect to the Globally Harmonized Classification system. It

has been shown that the probabilities of correct classification is greatest when test doses and category limits are identical. The minimal distance factor between two neighboring toxic classes has to be 4 for slopes of at least 1 to achieve a probability of correct classification of at least 0.5 for at least one LD_{50} value in each category. For a slope of at least 1 the probability of an allocation to a lower than correct toxic category is limited to 0.256.

14. There is only a low dependence on the starting dose with respect to classification results, especially for slopes of greater than 1. With increasing slopes or increasing LD_{50} values this influence decreases and tends toward zero for an unlimited increase of slope or LD_{50}. Also for infinitely low values of LD_{50} the influence becomes zero.

15. There is a strong dependence on the starting dose with respect to expected numbers of animals used and of moribund/dead animals. Therefore an appropriate starting dose should be near the true LD_{50} of the substance to be tested to minimise the number of animals used.

Point estimate of LD_{50}

16. The ATC was not designed to determine a point estimate of LD_{50}. However, a point estimate of the LD_{50} can be calculated by the maximum likelihood method providing there are at least two doses with mortality rates not equal to 0% or 100%. However, the probability of two such doses is rather low because the distance between two neighboring doses is 6- to 10-fold and up to six animals per dose are used (10).

Confidence Limits On The Estimate Of LD_{50}

17. The ATC was not designed to determine a point estimate of LD_{50}, or confidence limits. Providing there are at least three doses, two of which have mortality rates not equal to 0% or 100%, the maximum likelihood method can be used to calculate and broad confidence limits on the estimated LD_{50}.

Dose-Effect Curve

18. The ATC was not designed to determine a dose-effect curve for the LD_{50}. However, dose-effect curves can be calculated by the maximum likelihood method providing there are at least three doses, two with the specific toxic signs not present in 0% or 100% of the animals.

GUIDELINE 425:UP-AND-DOWN METHOD

Principles Underlying the Test Method

19. The concept of the up-and-down (UDP) testing approach (sometimes called a Staircase Design) was first described by Dixon and Mood (11)(12). There have been papers on such issues as its use with small samples (13) and its use with multiple animals per dose (14). One of the most extensive discussions appears in a draft monograph prepared by W. Dixon and Dixon Statistical Associates for a U.S. National Institutes of Health [NIH] Phase I Final Report, Reduction in Vertebrate Animal Use in Research, produced under SBIR Grant No. 1-R43-RR06151-01(15). This draft monograph is available from its author for a fee or from the National Center for Research Resources of the NIH to individuals under the Freedom of Information Act.

20. In 1985, Bruce proposed the use of the UDP for the determination of acute toxicity of chemicals (16). While there exist several variations of the up-and-down experimental design, Guideline 425 is a modification of the procedure of Bruce as adopted by ASTM in 1987 (17). The guideline provides a main

test, for LD_{50} point estimation and a computational procedure, used together with the main test to calculate confidence intervals. The UDP calls for dosing individual animals of a single sex, usually females, in sequence at 48-hour intervals, with the initial dose set just below "the toxicologist's best estimate of the LD_{50}," or at 175 mg/kg if no such estimate is possible. Following each death (or moribund state) the dose is lowered; following each survival, it is increased, according to a pre-specified dose progression factor. If a death follows an initial direction of increasing doses, or a survival follows an initial direction of decreasing dose, additional animals are tested following the same dose adjustment pattern and testing is ended if certain criteria are met. The OECD 425 protocol calls for a default dose progression factor of 3.2 and default s for maximum likelihood calculations of 0.5 (i.e., log(3.2)). Dosing levels and calculation details are provided in the guideline.

Point Estimate of the LD_{50}

21. From the data a point estimate of the LD_{50} is calculated using the maximum likelihood method (18)(19).

Confidence Limits On The Estimate Of LD_{50}

22. Confidence limits around the LD_{50} value can be calculated using the maximum likelihood method (18)(19), provided a suitable historical or other sound estimate of the standard deviation can be employed. A computational procedure based on profile likelihoods can provide confidence limits for the LD_{50} when no prior estimate of the standard deviation is available. The procedure identifies bounds for LD_{50} from a ratio of likelihood functions optimized over *sigma* (profile likelihoods). Procedures are also included for certain circumstances where no intermediate doses exist (for instance, when testing has proceeded through a wide range of doses with no reversal or where doses are so widely spaced that each animal provides a reversal).

Dose-Effect Curve

23. A dose effect curve can be calculated using a two parameter probit model provided that the response is quantal and there is an overlapping of the range of doses that result in a positive and negative response.

REFERENCES TO ANNEX 3

(1) Van den Heuvel, M J, D G Clark, R J Fielder, P P Koundakjian, G J A Oliver, D Pelling, N J Tomlinson, and A P Walker (1990). The international validation of a fixed-dose procedure as an alternative to the classical LD50 test. Fd. Chem. Toxicol., 28:469-482.

(2) British Toxicology Society (1984). Special report: a new approach to the classification of substances and preparations on the basis of their acute toxicity. Human Toxicol., 3:85-92.

(3) Van den Heuval, M J, A D Dayan and R O Shillaker (1987). Evaluation of the BTS approach to the testing of substances and preparations for their acute toxicity. Human Toxicol., 6:279-291.

(4) Lipnick, R L, J A Cotruvo, R N Hill, R D Bruce, K A Stitzel, A P Walker, I Chu, M Goddard, L Segal, J A Springer, and R C Myers (1995). Comparison of the Up-and-Down, Conventional LD50, and Fixed-Dose Acute Toxicity Procedures. Fd Chem. Toxic. 33: 223-231.

(5) Yam, J, P J Reer, and R D Bruce (1991). Comparison of the Up-and-Down Method and the

Fixed-Dose Procedure for Acute Oral Toxicity Testing. Fd Chem. Toxic. 29:259-263.

(6) Stallard N and A Whitehead (1995). Reducing numbers in the fixed-dose procedure. Human Expt. Toxicol., 14:315-323.

(7) Stallard N, A Whitehead and P Ridgway (2000). Statistical evaluation of modifications to the fixed dose procedure (manuscript in preparation).

(8) Diener, W; U Mischke, E Schlede and D Kayser (1995). The biometric evaluation of the OECD modified version of the acute toxic category method (oral). Arch.Toxicol. 69: 729-734.

(9) Schlede, E; U Mischke, W Diener and D Kayser (1995). The international validation study of the acute-toxic-category method (oral). Arch.Toxicol. 69: 659-670.

(10) Diener, W; L Siccha, U Mischke, D Kayser and E Schlede (1994). The biometric evaluation of the acute-toxic-category method (oral). Arch.Toxicol. 68: 599-610.

(11) Dixon, W J, and A M Mood (1948). A method for obtaining and analyzing sensitivity data. J. Amer. Statist. Assoc. 43:109-126.

(12) Dixon, W J (1991). Staircase Bioassay: The Up-and-Down Method. Neurosci. Biobehav. Rev. 15:47-50.

(13) Brownlee, K A , J L Hodges, Jr., and M Rosenblatt (1953). J Amer. Stat. Assoc., 48:262-277.

(14) Hsi, B P (1969). J Amer. Stat. Assoc., 64:147-162.

(15) Dixon, W J (1991). Design and Analysis of Quantal Dose-Response Experiments (with Emphasis on Staircase Designs). Dixon Statistical Associates, Los Angeles CA, USA.[discussion paper prepared for a US NIH Phase 1 Final Report: Reduction in Vertabrate Animal Use in Research , produced under SBIR Grant N° 1 R43-RR06151-01].

(16) Bruce, R D (1985). An up-and-down procedure for acute toxicity testing. Fundam. Appl. Tox., 5:151-157.

(17) ASTM (1987) E 1163-87, Standard test method for estimating acute oral toxicity in rats. American Society for Testing Materials, Philadelphia PA, USA.

(18) Dixon, W J (1965). The up-and down method for small samples. J. Amer. Statist. Assoc., 60:967-978.

(19) Finney, D J (1971). Probit Analysis, 3rd ed., Cambridge University Press, Cambridge, England, 50-80.