udpfin7
10/31/01

Acute Oral Toxicity: Up-and-Down Procedure

INTRODUCTION

1. 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 the American Society for Testing and Materials (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).

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 (CI), the test allows the observation of signs of toxicity. This test does not provide information about the slope of the dose-response curve. Appendix A contains definitions of some terms used in this guideline. Revision of this test guideline was undertaken concurrently with two other alternatives to conventional acute oral toxicity testing. An international guidance document on acute toxicity gives more information (12).

3. The guideline significantly reduces the number of animals used in comparison to Guideline 401, which often required at least 20 animals in a test: 1) the stopping rule limits the number of animals in a test; 2) sequential dosing introduces further efficiencies in animal use; 3) initial dosing is now set to be below the LD50 increasing the percentage of animals in which dosing levels will be sub lethal and thereby providing some reduction in pain and distress; and 4) the use of a single sex reduces the number of animals needed and minimizes the variability in the test population. Importantly, the guideline contains a requirement to follow the Organization for Economic Cooperation and Development (OECD) Guidance Document on Humane Endpoints (13) that should reduce the overall suffering of animals used in this type of toxicity test.

INITIAL CONSIDERATIONS

4. All available information on the test substance should be considered by the testing laboratory prior to conducting the study. Such 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. This information is useful to determine the relevance of the test for the protection of human health and the environment. This information may be valuable in selecting a dose other than the default starting dose. (See caveats in paragraph 5 about placement of starting dose.) For example, data from *in vitro* cytotoxicity assays can be useful as one of the tools in setting a starting dose for the *in vivo* assessment of acute oral toxicity (14, 15, 16). A Guidance Document on Using *In Vitro* Data to Estimate *In Vivo* Starting Doses for Acute Toxicity is available (15), and preliminary information suggests that the use of this approach may further reduce the number of animals used for *in vivo* testing (16).

5. If 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 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 paragraph 30 and Appendix B 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 31.)

6. 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.

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

8. 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).

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

PRINCIPLE OF THE LIMIT TEST

10. 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 22 - 24 for limit test at 2000 mg/kg and paragraphs 25 - 28 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

11. 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 LD50. 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. Paragraph 30 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 29 and 33 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 32). Dosing is stopped when one of these criteria is satisfied (see paragraphs 31 and 39), at which time an estimate of the LD50 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 LD50 is calculated using the method of maximum likelihood (17)(18). (See paragraphs 39 and 41.)

12. 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 43.

DESCRIPTION OF THE METHOD

Selection of animals species

13. The preferred rodent species is the rat although other rodent species may be used. 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 should be tested (12). 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 (7). 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.

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.

14. 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 ± 20 % of the mean initial weight of all previously dosed animals.

Housing and feeding conditions

15. 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.

Preparation of animals

16. 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.

Preparation of doses

17. 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.

18. Constant concentration should be used in dosing unless there is clear scientific or regulatory justification for not dosing so. In the event that constant volume was used, information on the actual concentrations used should be provided. 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.

Administration of doses

19. 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.

20. 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.

PROCEDURE

Limit test and main test

21. 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.

Limit test

Limit test at 2000 mg/kg

22. 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. The second and third animals can be dosed concurrently and the fourth and fifth sequentially. However, if three animals die, the limit test is terminated and the main test is performed. 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 29 for initial observation period). Late deaths should be counted the same as other deaths. The results are evaluated as follows (O=survival, X=death).

23. The LD50 is less than the test dose (2000 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

24. 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 OO/X
O XX OO

Limit Test at 5000 mg/kg

25. Exceptionally, and only when justified by specific regulatory needs, the use of a dose at 5000 mg/kg may be considered (see Appendix D). Recognizing the need to protect animal welfare, testing of animals in class 5 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 or the environment.

26. 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). 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).

27. 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 XU (U can be O or X, the dosing of the 5th animal is not necessary)

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

O OO UU (U can be O or X, the dosing of the 4th and 5th animal is not necessary)
O XO XO
O XO OU (U can be O or X, the dosing of the 5th animal is not necessary)
O OX XO
O OX OU (U can be O or X, the dosing of the 5th animal is not necessary)
O XX OO

Main test

29. 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.

30. The first animal is dosed a step below the toxicologist's best 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. 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, 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. (Appendix B includes a table of dose progressions for whole number slopes ranging from 1 to 8 with starting dose 175 mg/kg.)

31. 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 42 and Appendix C. 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.

32. 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 38 and 39.

33. 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 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 LD50 down.

Observations

34. 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 (19). All observations are systematically recorded with individual records being maintained for each animal.

35. 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 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 (13) 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.

Body weight

36. 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

37. 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

38. 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 (19), 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

39. The LD50 is calculated using the maximum likelihood method (17)(18)**,** except in the exceptional cases described in paragraph 40. 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 ith animal survived, or
$L_i = F(Z_i)$ if the ith animal died,

where

F = cumulative standard normal distribution,
$Z_i = [\log(d_i) - mu\ ] / sigma$
d_i = dose given to the ith 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 LD50 is given by the value of *mu* that maximizes the likelihood L (see paragraph 41).

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

40. 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 31 (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.

(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 31 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.

41. Maximum likelihood calculation can be performed using either SAS (17) (e.g., PROC NLIN) or BMDP (18) (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 the Up-and-Down Procedure.) The program's output is an estimate of log(LD50) and its standard error.

42. The likelihood-ratio stopping rule (c) in paragraph 31 is based on three measures of test progress, that are of the form of the likelihood in paragraph 39, 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 31. The equations for the likelihood-ratio criteria are provided in Appendix C. 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 Appendix C. If the criterion is met, testing stops and the LD50 can be calculated by the maximum likelihood method.

<u>Computation of confidence interval</u>

(43) 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. In this case, 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. In this case, 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 is expected to be close to the original estimated LD50 and both of these estimates are expected to be close to the true LD50.

Depending on the outcome of the main test, one of two different types of interval estimates of the

true LD50 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 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 (20). 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 (21).

(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 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.

In some instances, confidence intervals 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.

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 (22) also available through the EPA. Appendix E provides information on choice of dose progression and initial dose level for the UDP and describes test performance under a variety of circumstances.

Report

44. The test report must include the following information:

Test substance:

- physical nature, purity and physicochemical properties (including isomerization);
- identification data.

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;
- rationale for use of males instead of females;
- source, housing conditions, diet, etc.;
- individual weights of animals at the start of the test, at day 7, and at day 14.

Test conditions:

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

Results:

- body weight/body weight changes;
- 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);
- 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. (1965). The Up-and-Down Method for Small Samples. 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 LD50 and Fixed Dose Acute Toxicity Procedures. Fd. Chem. Toxicol., 33, 223-231.

(8) Choi, S.C. (1990). Interval Estimation of the LD50 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). Revised Draft Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation.

(14) 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.

(15) 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.

(16) Spielmann, H.E., M.Genschow, M. Leibsch, and W. Halle. (1999). Determination of the Starting Dose for Acute Oral Toxicity (LD50) Testing in the Up-and-Down Procedure (UDP) from Cytotoxicity Data ATLA 27: 957-966.

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

(18) 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.

(19) 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.

(20) 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.

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

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

APPENDIX A

DEFINITIONS

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 is an interval estimate, a range of values, intended to include the true LD50 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) 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).

LD50 (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 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 (2000-5000 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).
Probity is an abbreviation for the term "probability integral transformation" and a probity 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 probity.

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 probity analysis, when responses are analyzed on a probity 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 probity 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 5 as a shorthand for the criterion that relies on comparison of ratios to a critical value.

APPENDIX B

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 5 of the test 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 Up-and-Down Procedure 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 B.1 provides dose progressions for whole number multiples of slope, from 1 to 8.

Table B.1. Dose Progressions for Up-and-Down Procedure

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*
							0.243	0.233
					0.28	0.26		
				0.31			0.34	0.31
			0.38			0.38		
								0.41
					0.44		0.47	
		0.55		0.55		0.55		0.55
					0.70		0.65	
								0.74
			0.81			0.81		
				0.98			0.91	0.98
					1.10	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

Table B.1. (continued)

Slope =	1	2	3	4	5	6	7	8
			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

APPENDIX C

COMPUTATIONS FOR THE LIKELIHOOD-RATIO STOPPING RULE

As described in Guideline paragraph 31, 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 are satisfied, dosing would stop when 15 animals are dosed. Tables C.1 - C.4 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.

Table C.1 shows how the main test would stop if 3 animals have survived at the limit dose of 2000 mg/kg; Table C.2 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 C.3 shows how a particular sequence of 5 reversals in 6 tested animals could occur and allow test completion. Finally, Table C.4 illustrates a situation several animals into a test, 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.

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 CI by a likelihood-based procedure.

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.

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.

The calculations are easily performed in any spreadsheet with normal probability functions. The calculations are illustrated in Table C.4, 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 C.4)
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 ith 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.

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.

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. <u>Likelihood for the rough LD50 estimate.</u>

"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.

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 39). The likelihood contribution for the ith animal is denoted L_i.

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 probity 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 ith 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 $F(1.64) \approx 0.95$.

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 ith 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. <u>Calculate likelihoods for two dose values above and below the rough estimate.</u>

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. <u>Calculate likelihood-ratios.</u>

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.

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, Millennium Edition. Cambridge, MA, USA.

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

Table C.1. 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. 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	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 C.2. 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				-						
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 C.3. 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				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 C.4. 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		
Calculated maximum likelihood estimate of LD50 =						1329.6						

Final estimate obtained from Maximum Likelihood Calculations

APPENDIX D

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

Test substances could be classified in the hazard classification defined by:
2000 mg/kg<LD50<5000 mg/kg (Class 5 in the Globally Harmonized System (GHS)) in the following cases:

a) if reliable evidence is already available that indicates that LD50 to be in the range of class 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 class is not warranted, and
   - reliable information is available indicating significant toxic effects in humans, or
   - any mortality is observed when tested up to class 4 values by the oral route, or
   - where expert judgement confirms significant clinical signs of toxicity, when tested up to class 4 values, except for diarrhea, piloerection or an ungroomed appearance, or
   - where expert judgement confirms reliable information indicating the potential for significant acute effect from the other animal studies.

APPENDIX E

PERFORMANCE OF THE UP-AND-DOWN PROCEDURE

This appendix 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 E.1).

1. Adjusting the Dose Progression and Initial Dose.

For optimum performance of the UDP, the dose progression used should be based on an accurate 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 LD50 estimate, it is essential that dosing be initiated below the actual LD50.

(i) 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 LD50 may be "biased" in the direction of starting dose. For example, if the starting dose is less than the true LD50 of the test population, the estimated LD50 will generally be below the true LD50. Also, if the starting dose is greater than the true LD50 of the test population, the estimated LD50 will tend to be greater than the true LD50. To minimize the chance of overestimating the LD50 due to this bias, the UDP guideline recommends a choice of starting dose just below the assumed LD50.

(ii) 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 LD50 can be much larger or much smaller than the true LD50 depending on the starting dose. In this case, the LD50 can be estimated only within a range. (This is Case 3 described below.)

2. Confidence Interval.

Coverage of the confidence interval is the probability that a calculated confidence interval encloses the true LD50 for an experimental sample. Because the profile likelihood method is approximate, coverage of the confidence interval 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 confidence interval is limited by the dose progression chosen. Generally, no type of confidence interval would be more narrow than the dose progression.

3. 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 - Case 5 below, 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. In considering these cases, note that doses can be repeated more than once in the course of sequential dosing.

Table E.1. Outcomes of the Up-and-Down Procedure: 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.	LD50 cannot be calculated. Confidence interval not applicable.	Possible inferences: 1a) LD50 < lowest dose; 1b) LD50 > 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 confidence interval are straightforward.	The LD50 can be estimated and its confidence interval 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 LD50 falls between the two bounding doses. Any value of LD50 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 LD50 is set at the single dose showing partial response and its confidence interval is calculated using profile likelihood method.	The LD50 can be estimated and its confidence interval 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 LD50 is set at the dose with the partial response. A profile likelihood confidence interval is calculated and may be finite or infinite.	The true LD50 could be at the boundary of the testing range with more or less confidence.