Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development

23-Jul-2001

English - Or. English

ENV/JM/MONO(2001)4
Unclassified

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

OECD SERIES ON TESTING AND ASSESSMENT
Number 24

GUIDANCE DOCUMENT ON ACUTE ORAL TOXICITY TESTING

JT00111082

English - Or. English

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.

OECD Environment Directorate,
Environment, Health and Safety Division

2 rue André-Pascal
75775 Paris Cedex 16
France

Fax: (33-1) 45 24 16 75

E-mail: ehscont@oecd.org

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.

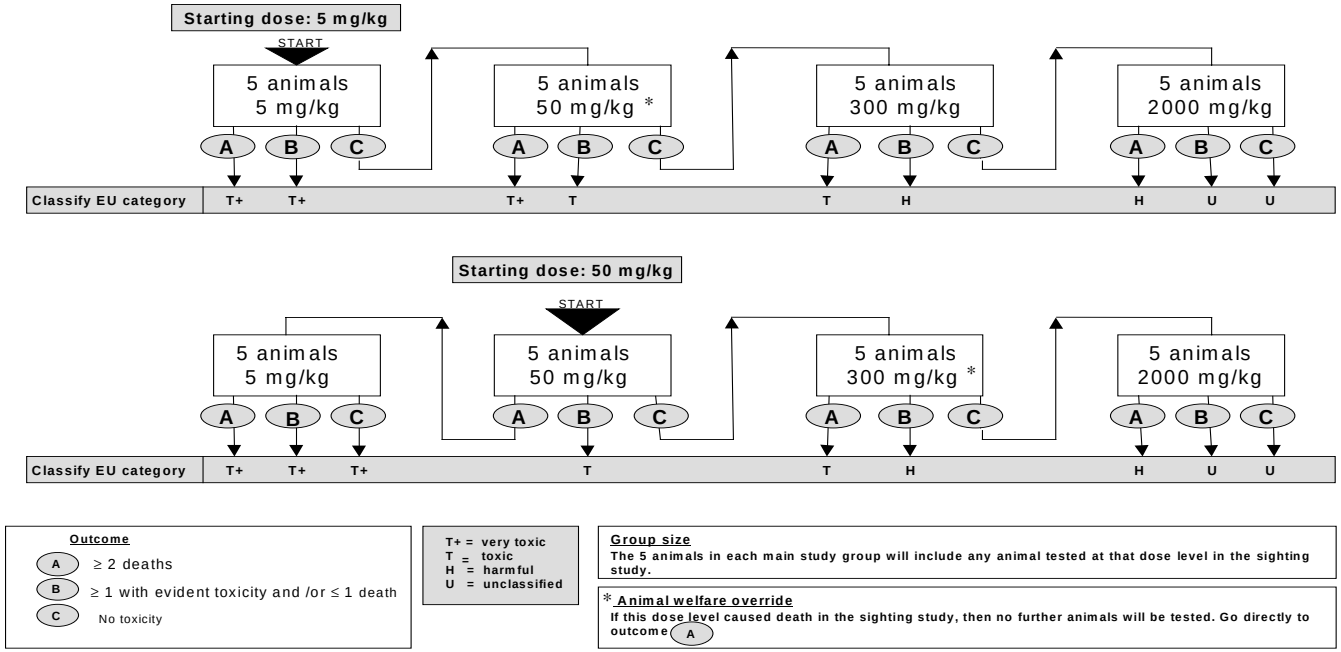
(6) OECD (2000). Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation. Environmental Health and Safety Monograph Series on Testing and Assessment No. 19, September 2000. [http://www.oecd.org/ehs/test/monos/htm]

(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.

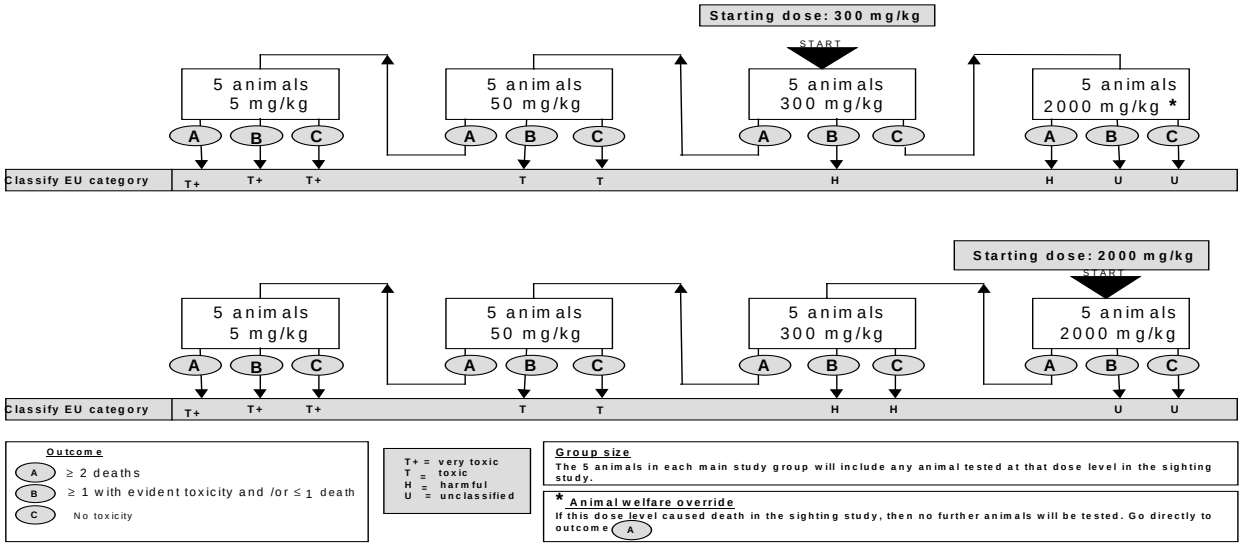
ANNEX 1

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)



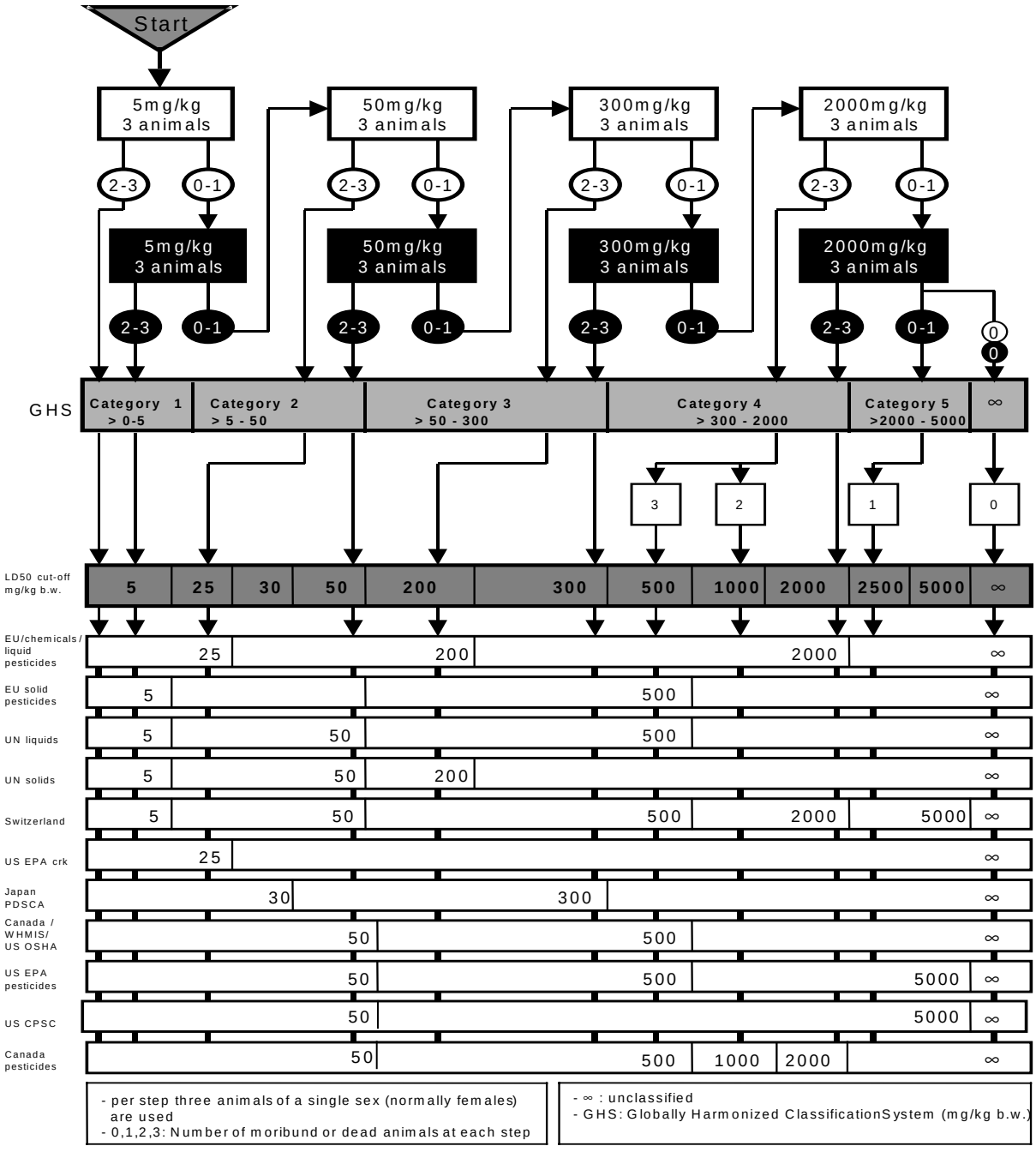
ANNEX 1 (continued)

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)



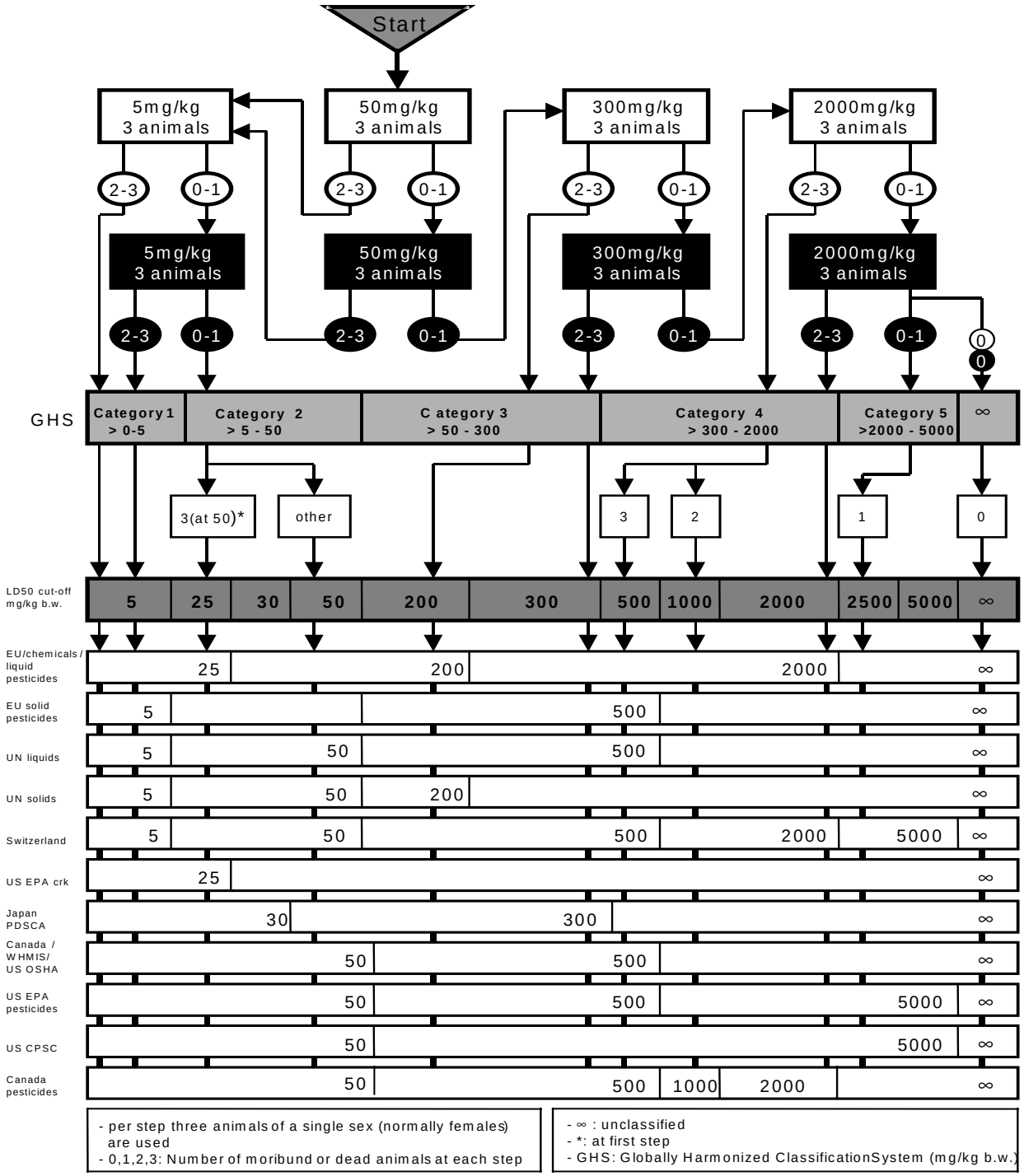
ANNEX 2

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)



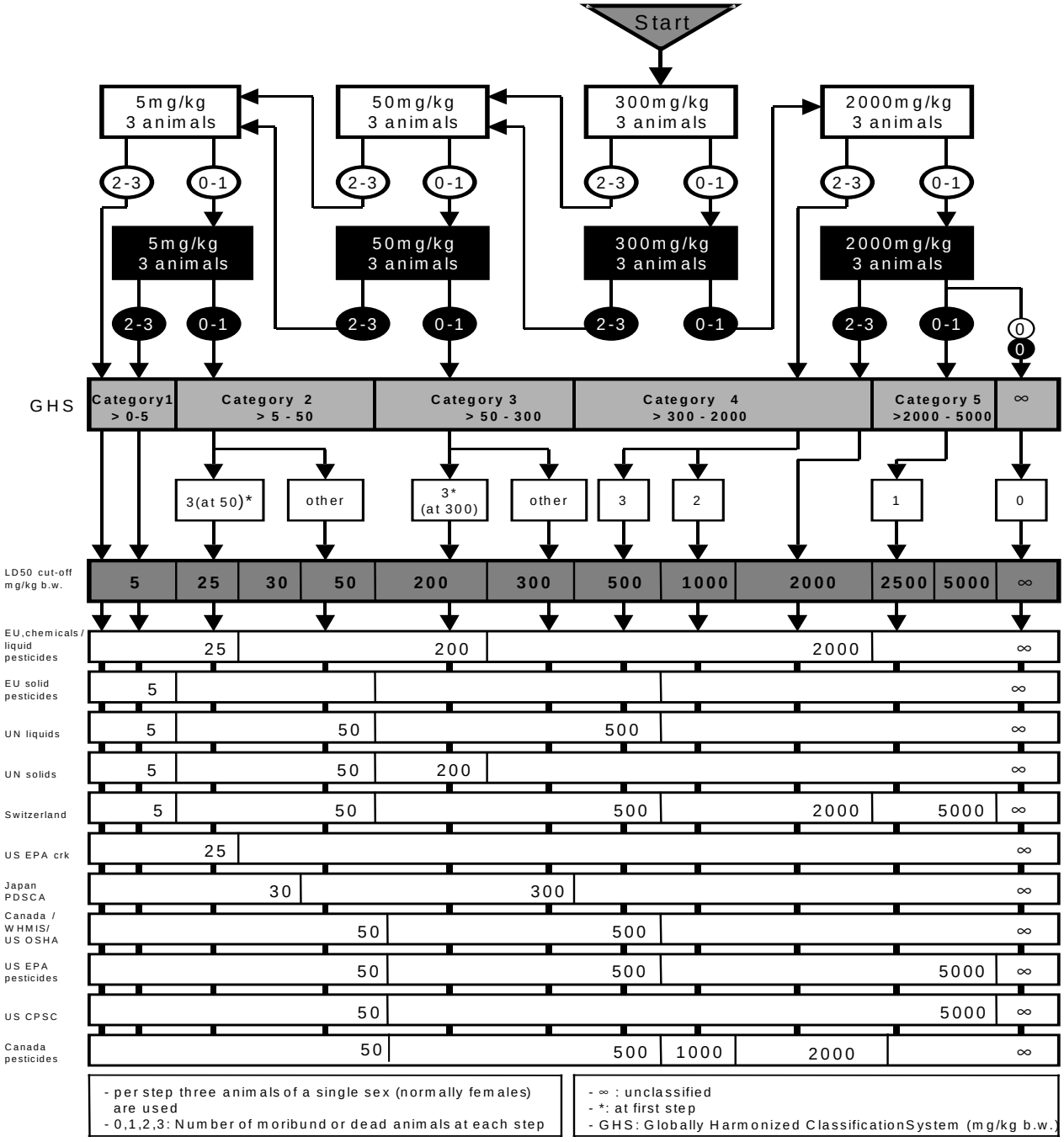
ANNEX 2 (continued 1)

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)



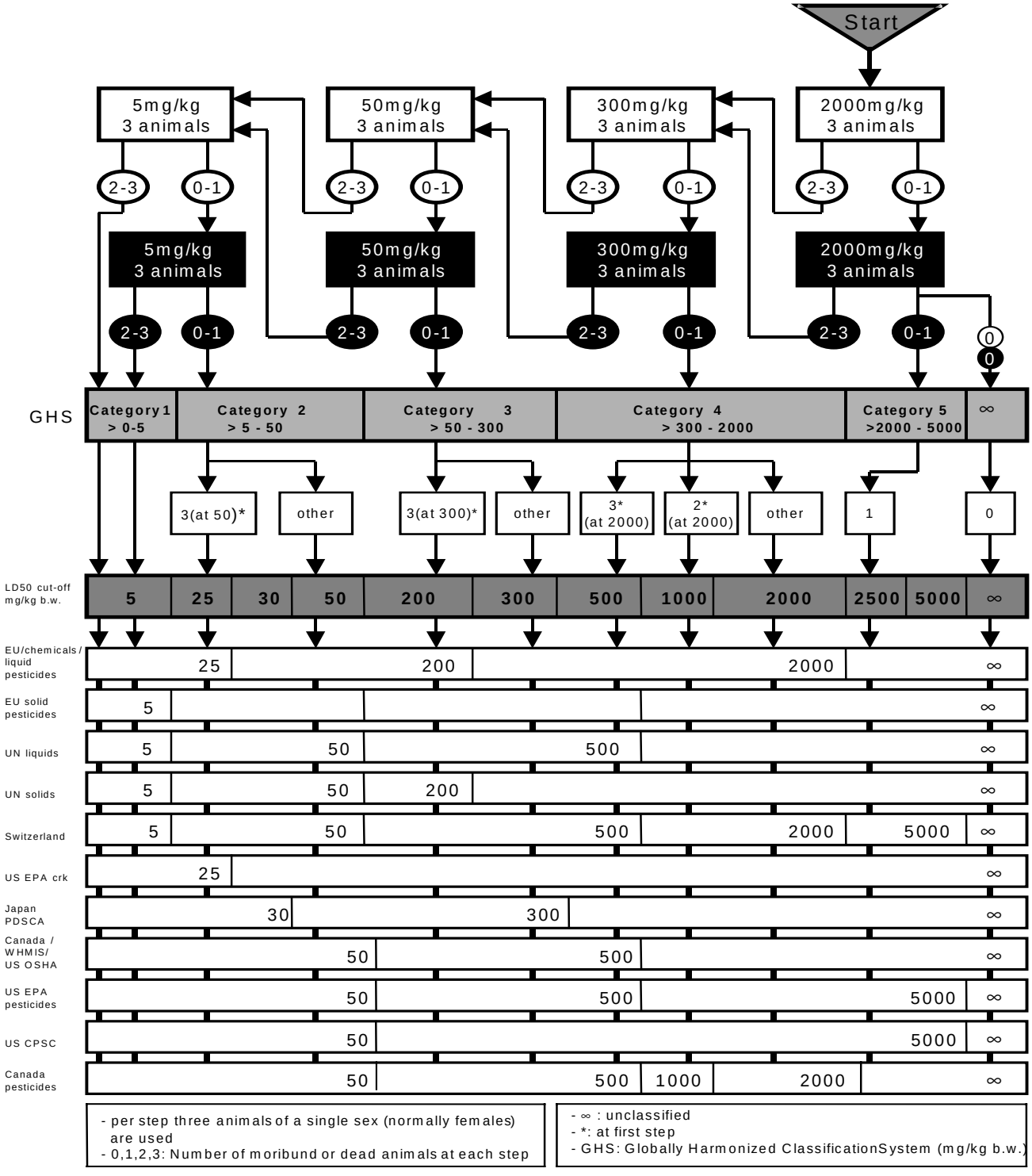
ANNEX 2 (continued 2)

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.