APPENDIX F

Up-and-Down Procedure
Revised Background Review Document (BRD)

October 31, 2001

Note: The April 2000 Background Review Document (BRD) was reviewed by the Peer Review Panel at the July 25, 2000 Panel meeting. This document was subsequently revised in accordance with the Panel's discussions, recommendations, and conclusions. To maintain continuity between the two BRDs, the designation for each appendix cited in the original BRD as well as the designation used in the UDP Peer Panel Report are provided. This revised BRD does not include information provided to the UDP Peer Panel for their August 21, 2001 deliberations.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Introduction: The acute oral toxicity test is a fundamental component in defining the toxicity of a test material for hazard classification and labeling purposes. There are two types of acute oral tests: a) those that identify a dose range in which the median lethal dose (LD50) falls, and b) those that determine a point estimate of the median lethal dose of the material. In tests that estimate the LD50, if sufficient data are available, an estimate of the slope of the dose-response curve and confidence interval can also be determined. In 1981, the Organization of Economic Co-operation and Development (OECD) adopted a test guideline (TG 401) for acute oral toxicity that estimated the LD50 and in many cases, the slope and confidence interval. TG 401 has become the traditional acute oral toxicity test. TG 401 was revised in 1987 to utilize three dose groups of five rats of one sex with confirmation in the other sex using one group of five rats. This resulted in reduced animal use from 50 or more in the 1981 version to 20 in the 1987 version.

Since 1987, OECD has adopted three additional acute oral toxicity tests, one of which is the up-and-down procedure (UDP) in 1998. With the new test guidelines adopted, OECD is considering a proposal to delete TG 401. Of the three alternative tests, the UDP is the only test providing a point estimate of the LD50 and does this rather efficiently for many chemicals by only using six or seven animals. However, the UDP does not provide an estimate of the slope of the dose-response curve and confidence interval. With TG 401 to be deleted, there would be no method available to regulatory agencies that provided an estimate of slope and confidence interval. In addition, the global harmonization of the classification scheme has resulted in the need to revise the Fixed-Dose Procedure (FDP) and the Acute Toxic Class Method (ATCM). As a result, OECD agreed to revise all three alternative methods. The U.S. Environmental Protection Agency (EPA) agreed to revise the UDP to include a procedure that would provide slope and corresponding confidence interval estimates. The UDP described in this document has been revised to include: a) a modified up-and-down procedure with improved performance; b) a modified Limit Test utilizing only females and providing a limit dose of 5000 mg/kg for specific regulatory purposes; and c) an added supplemental test for determining the slope and confidence interval.

Test Method Protocol: The Revised UDP has three tests: a) the primary test to estimate the LD50; b) a Limit Test allowing testing at 5000 mg/kg for specific regulatory purposes; and c) the added supplemental test to estimate the slope and confidence interval. In the primary test, one animal is dosed at 175 mg/kg and observed for 14 days. If the animal is alive at 48 hours, a second animal is dosed at a 0.5 log higher dose. If the first animal dies, then the second animal is dosed at a 0.5 log lower dose. Dosing stops when the stopping criteria are satisfied. In the Limit Test, one animal is dosed at 2000/5000 mg/kg. If the animal dies, the primary test is conducted. If the animal lives, two more are dosed at the limit dose. If they both live, the Limit Test is satisfied because three animals have survived at the limit dose. If one or both of the two animals die, then two more are tested at the limit dose. If a total of three animals live, the Limit Test is satisfied. If three animals die, the primary test is conducted. In the supplemental test, three up and down tests (runs) are started at slightly differing doses below the LD50. Dosing continues in each run until an animal dies.

Characterization of the Materials Used: There have been three validation studies of the UDP. A total of 25 chemicals were tested in which data using the UDP were compared to data generated using TG 401. A wide variety of chemicals from a number of chemical classes were tested, which affected different target organs and exhibited a wide range of LD50s (ranging from 48 to greater than 20,000 mg/kg).

Reference Data: Reference data consisted of acute oral toxicity data generated using TG 401. In two of the studies, the data for TG 401 and the UDP were generated concurrently in the same laboratory. In the third study, the chemicals were selected from published data from a validation study of ATCM. The data were generated in compliance with national or international GLP guidelines.

In Vivo Test Method Data and Results: Although the UDP was not adopted at the time, the protocol used a default starting dose of 100 mg/kg, a dose spacing factor of 1.3, and a stopping rule of testing four animals after the first reversal.

Computer Simulation Validation of Revised UDP: A statistical procedure involving 1000 to 5000 computer simulations examined many permutations of testing conditions and the range of results provided insight into the factors affecting the slope. These simulations allowed the determination of the recommended starting dose, the dose spacing factor, and the stopping rules.

In Vivo Test Method Performance Assessment: For the three validation studies, the absolute ratio of the LD50 from TG 401 studies to the LD50 from UDP studies average 1.76, well within expected variability. If one apparent outlier is eliminated, the ratio becomes 1.28. The one exception was for mercuric chloride.

Computer Simulation Performance Assessment: Simulations have resulted in changing the starting dose, the dose spacing factor, and stopping rules. The default starting dose was increased from 100 mg/kg to 175 mg/kg as a compromise between the possibility of severe toxicity and starting too far from the LD50. The dose spacing factor was changed to 3.2 to allow the investigator to move more quickly toward the LD50 if the starting dose was far from the LD50 and to better estimate the LD50 for chemicals with a shallow slope. The stopping criteria include maximum likelihood ratios and allow a more accurate estimate of the LD50 without utilizing too many animals.

Test Method Reliability: There are no known *in vivo* data on the reliability of the Revised UDP. A number of inter- and intra-laboratory validation studies were conducted prior to 1981. Considering the extremes in testing conditions, it is remarkable that the LD50 varied by no more than a factor of 2 to 3. These studies showed the need to standardize the protocol for toxicity methods. Under standardized protocols, the variability was greatly reduced. In the three validation studies, the absolute ratio of the LD50 for the UDP data and TG 401 data was 1.76. When mercuric chloride was not considered, the ratio was 1.28. These ratios are well within the expected reliability factor of three.

Test Method Data Quality: The data for the three validation studies were generated under applicable GLPs and no discrepancies were noted that altered the general conclusions of the study reports.

Other Scientific Reports and Reviews: No other published UDP data in mammals are available. Unpublished data in birds dosed two at a time results in using large numbers of animals. Consideration was given to the moving-average method for estimating the slope and confidence interval.

Animal Welfare Considerations: There was a clear reduction in incidence of pain and suffering in animals in the UDP study compared to TG 401 animals. The UDP reduced animal usage by 77% compared to animal usage in TG 401 studies. The Revised UDP emphasizes the utilization of humane endpoints and the handling of moribund animals. Although it has been suggested that cytotoxicity tests replace acute oral testing in animals, *in vitro* cytotoxicity tests have not been validated as replacement tests.

Other Practical Considerations: Gender differential sensitivity, equipment, and training were addressed. Based on studies that display sex differences in sensitivity, the female is considered more sensitivity and will be used except when known male sensitivity dictates otherwise. To conduct Revised UDP studies, laboratories will need a computer and access to readily available commercial software. Software may be made available on the OECD and EPA websites. The technical staff will need to be familiar with humane endpoints and the handling of moribund animals. In addition, they will need to be able to use the computer to conduct the studies properly to evaluate stopping rule criteria as well as the LD50 and slope

estimates. The Revised UDP will take at least two weeks to complete dosing and therefore at least four weeks to complete the study. Although there will be fewer animals to observe at any given time, the cost of the study may increase because of the extended time to conduct the study.

1.0 Introduction and Rationale of the Revised UDP

1.1 Introduction

1.1.1 Human Poisonings

Acute exposure to poisonous substances is a common occurrence. For example, in the United States, based on data for 1998 from the Toxic Exposure Surveillance System (65 Poison Control Centers serving 257.5 million people), a total of 2,241,082 human exposures were reported resulting in 8.7 exposures per 1000 people. Of these exposures, 775 fatalities were reported with the highest incidence (432, 56%) in persons between 20 and 49 years of age. Of these totals, 1,749,792 exposures (78%) and 638 fatalities (82%) were via oral ingestion. Of the total exposures, 86,289 (3.9%) were to pesticides while the highest incidence of exposure was to cleaning substances (229,500; 10.2%). Insecticides accounted for only 16 deaths (2.1%) compared to 246 (32%) following ingestion of analgesics.

1.1.2 Acute Toxicity Testing

The purpose of acute toxicity testing is to identify and categorize those chemical substance (hereafter referred to as substances) that pose a potential hazard to humans and other species. Historically, in determining the acute toxicity of a substance, one of the first tests to be conducted has been an acute oral toxicity test designed to estimate an acute oral LD50. The LD50, or median lethal dose, is the dose expected to kill 50% of the test population. The test animal of choice for acute lethality testing has been the rat, although acute oral LD50 values have been calculated for mice and other mammalian species. Birds, fish, and other species have been used for ecological considerations. The classical method for estimating the LD50 has been to orally dose individual animals, in groups of five to ten per sex, with varying concentrations of the test substance and to subsequently observe whether the animal lived or died over a defined period of time (generally 14 days). The calculation of the LD50 is derived from the dose-response curve for lethality. The confidence limits of the LD50 and an estimate of the slope of the dose-response curve can be calculated under two conditions: (1) when there are at least two doses in which at least one, but not all, of the animals are killed, or (2) if the dose range for surviving animals overlaps sufficiently the dose range for animals that die.[1]

A procedure for calculating the oral LD50 was first described by Trevan in 1927. This approach has been used as a benchmark for comparing the acute toxicity of substances and relating their toxicity to human health. Inspection of oral LD50 data in large databases (e.g., the Registry of Toxic Effects of Chemical Substances [RTECS], the International Uniform ChemicaL Information Database [IUCLID]) indicates that multiple values obtained for the same test substance in the same species can be quite variable. However, much of these data were generated using experimental conditions varying widely with respect to strain, sex, age, husbandry, and health status of the animals. As regulatory agencies began to require

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

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

Further, the U.S. EPA defines *sigma* as 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.

acute oral toxicity data, it became evident that a standardized protocol(s) must be used if data for test substances are to be valid and useful.

The U.S. Environmental Protection Agency (EPA) published test guidelines for acute toxicity in October 1982 as part of Subdivision F of the Pesticide Assessment Guidelines for the Office of Pesticides and in September 1985 as part of 40 CFR part 797 for the Office of Toxic Substances. Since publication of the guidelines, the results of more than 15,000 acute oral toxicity tests have been submitted for consideration to the U.S. EPA's Office of Pesticides. Similarly, the Consumer Product Safety Commission (CPSC) utilizes acute oral toxicity in regulating commercial products in the United States (16 CFR Part 1500; original BRD **Appendix E**, currently **Appendix Q-1)**. In contrast, the Food and Drug Administration (FDA) does not require this type of acute toxicity testing for drugs.

1.1.3 The Traditional LD50 Test

The LD50 method was further standardized in 1981 by the international acceptance among the member countries of the Organisation for Economic Co-operation and Development (OECD) of Test Guideline (TG) 401. In this test, the test substance is typically administered by oral gavage to fasted young adult animals (five animals per sex). The guideline calls for a minimum of three dose levels in the toxic/lethal range; generally, however, the test typically included at least five dose levels to ensure adequate data for calculating an LD50. For test substances with no information regarding their potential for acute oral toxicity, a range-finding or sighting study of up to five animals could be conducted to identify the range of lethal doses. In such situations, at least 30 animals per sex are utilized in each test.

Generally, to minimize study duration and variation in dosing solutions, all dose groups are treated simultaneously. The animals are observed periodically during the first 24 hours with special attention given during the first four hours, then at least once a day for 14 days or until they recover. Clinical signs, including time of onset, duration, severity, and reversibility of toxic manifestations, are recorded at each observation period. Body weights are determined pre-treatment, weekly thereafter, and at the death of the animals or termination of the study. All surviving animals are humanely killed at 14 days or after recovery, whichever is earlier. Gross necropsies are conducted on all animals in the study. The goal of the test is to have at least two groups for each sex in which at least one, but not all, animals are killed by the test substance. If this circumstance occurs, the slope of the dose-response curve and confidence interval can be calculated using probit analysis. A Limit Test, which involves the dosing of five animals of each sex at 5000 mg/kg, is used for substances with low toxicity. If, for each ex, no more than two animals die, then the LD50 for that sex is considered to be greater than 5000 mg/kg. Variation in the results due to inter-animal variability, intra- and inter-laboratory variability, and to differences in strain, sex, estrus cycle, and species have been characterized. Based on intra- and inter-laboratory testing, the point estimate of the LD50 appears to be reliable within a factor of two or three (Griffith, 1964; Weil et al., 1966; Weil and Wright, 1967). If appropriate data are obtained, OECD TG 401 can provide the LD50, the slope, the confidence interval, and the hazard classification.

In 1987, in response to concerns about the numbers of animals used in LD50 testing, OECD TG 401 was revised to require only one sex with confirmation in the other sex at one dose level only (OECD, 1987) (original BRD **Appendix A,** final report **Appendix I).** This revision reduced the minimum number of animals required for each test from 50 to 60 to between 25 and 30. Also, in the 1987 version of OECD TG 401, the number of animals for the Limit Test was reduced to five animals of a single sex dosed at 2000 mg/kg.

Additional efforts have been made to reduce the number of animals used while maintaining the accuracy of the method for assessing the acute toxicity of a test substance. These alternative approaches **do not** involve a change in the treatment of the animals or in the endpoints examined. Since 1987, OECD has

approved three additional acute oral toxicity test guidelines that reduce animal use: TG 420 (the Fixed-Dose Procedure; FDP) in July 1992 (OECD, 1992); TG 423 (the Acute Toxic Class Method, ATCM) in March 1996 (OECD, 1996); and TG 425 (the UDP) in October 1998 (OECD, 1998). OECD TG 420 and TG 423 do not provide a point estimate of the LD50, but do provide a dose range in which the LD50 is expected to occur.

1.1.4 The UDP (OECD TG 425)

The UDP, a sequential test method, was first described by Bruce (1985). Three validation studies have been conducted to evaluate the ability of the UDP to estimate the LD50 when compared to the traditional LD50 method described in OECD TG 401 (Bruce, 1987; Bonnyns et al., 1988; Yam et al., 1991). Based on these studies and other considerations, in 1998, the OECD adopted the UDP (TG 425) as an acute oral toxicity test. The 1998 OECD TG 425 entitled "Acute Oral Toxicity: Up-And-Down Procedure" is provided in **Appendix H** of this final report (original BRD **Appendix A**).

In this test, one animal (usually a female) is dosed at the best estimate of the LD50, with 200 or 500 mg/kg suggested as a default-starting dose level if no toxicity information is available. If the animal dies or is moribund within 24 hours of dosing, a second animal is dosed at a lower dose level. If feasible, a dose-spacing factor of 1.3 is used, but other dose-spacing factors may be used if justified. If the first animal survives, a second animal is dosed at an appropriate higher dose level. Dosing continues until four animals are dosed after the first reversal (minimum of 6 animals). Information from one sex may be adequate to assess acute toxicity. However, if desirable, comparability of response in the other sex can be evaluated by administering to generally not more than three animals, dose levels around the estimated LD50. In the Limit Test, if the first animal dosed at 2000 mg/kg survives, the second animal is treated with the same dose level. When three animals have survived at the limit dose level, three animals of the opposite sex are dosed at the same dose level to verify the absence of acute toxicity. If all animals survive, then the LD50 is considered to be greater than 2000 mg/kg. The UDP employs a parameterized maximum likelihood method to estimate the LD50, which is used to identify the toxic class of the substance for labeling purposes (see U.S. EPA Document 4; original BRD **Appendix C**, final report **Appendix J-3**).

At the March 1999 OECD Expert Meeting (Washington, DC, U.S.), it was recognized that there were strengths and weakness in each of the acute oral toxicity tests (OECD TG 401, TG 420, TG 423, TG 425). Although acute toxicity information is used primarily to classify and label substances, some authorities also use acute toxicity test results to perform various risk assessment functions, including a determination of confidence interval and slope to make risk projections at the low end of the dose-response curve. Among the acute toxicity tests, only OECD TG 401 provided the ability to measure risk assessment parameters and OECD had decided to phase out this guideline. In recognition of the concerns identified at this meeting, it was decided that the alternative test guidelines to OECD TG 401 required revision. As part of the revision process, authorities revising the guidelines were charged with incorporating a number of considerations, including: (1) restricting the test to females only; (2) incorporating the new globally harmonized classification scheme (OECD, 1998); (3) adding an optional range-finding assay; (4) incorporating an ability to evaluate toxicity in the range of LD50 values of 2000 to 5000 mg/kg body weight; and (5) changing the test design to improve the operating characteristics of the method when the approximate LD50 is unknown or for substances with a low dose-response slope. In the case of OECD TG 425, the U.S. EPA was asked also to add a procedure for estimating the slope of the dose-response curve (the slope of the dose-response curve defines the confidence interval for the LD50) (see U.S. EPA Document 12; original BRD **Appendix C**, final report **Appendix O**). Other major motivations for revising the UDP were:

1. computer simulations had revealed that the UDP was biased towards the starting dose level for test substances with a shallow slope; and

2. the UDP could require significantly more animals per test if the starting dose level was far from the LD50.

Computer simulations were performed to evaluate the performance of the UDP as described in OECD TG 425 and to determine appropriate changes to optimize the method's performance without actually testing animals in the laboratory. Efforts to revise the UDP proceeded along two lines:

1. To revise the single-sequence version of the UDP to improve its performance when the approximate LD50 and dose-response slope are not known or for substances with wide variability of response, and to allow for lethality to be evaluated in the 2000 to 5000 mg/kg range for certain hazard classification purposes.
2. To provide a multi-sequence test method that can simultaneously address the issues in #1, while also providing the confidence interval and slope. This method would allow for both hazard classification and risk assessment needs.

1.1.5 The Regulatory Need for Slope and LD50 Confidence Intervals

The regulatory need for slope and confidence limits is based on the requirements of ecological risk assessment. In assessing the risk of pesticides to nontarget organisms, the U.S. EPA compares toxicity information with the expected environmental concentration and subsequently determines the likelihood that nontarget organisms will be exposed. When lethality is the toxic effect of concern, the results of acute toxicity testing are used. Laboratory data on the rat are used as surrogate information for naturally occurring populations of terrestrial animals. For assessment of hazard to other nontarget species, the U.S. EPA receives data on aquatic and avian species. Acute toxicity data used include the LD50 value, the slope of the dose-response curve, and information on dose effects. Risk assessment involves comparison of hazard and exposure to characterize risk. Risk assessments are performed to determine the existence of a population loss potential from the use of pesticides in the environment. In addition, the U.S. Endangered Species Act mandates that the U.S. EPA assess the potential for individual deaths of listed species due to use of pesticides.

1.1.5.1 Range of Data Available

Data available at the time of registration or reregistration of a pesticide consist of laboratory studies of toxicity and environmental fate. In addition, pesticide registrants submit small plot field studies of pesticide behavior in the environment. Effects in nontarget organisms are characterized primarily by using single-species laboratory toxicity tests, which yield dose-response curves of lethality and effect. This information can be augmented by data on effects of the substance in other nontarget species. Exposure estimates can be based on laboratory studies and any available monitoring data. Computer modeling can be used to generate distributions of expected environmental concentrations.

1.1.5.2 Use of Point Estimates

Preliminary risk assessments involve comparison of point estimates of toxic effects with point estimates of exposure (i.e., the most probable expected exposure). For acute toxicity to terrestrial vertebrates, for example, the expected environmental exposure can be compared at 20% of the LD50 as a regulatory threshold. The value of 20% LD50 has been traditionally used to initiate regulatory action in the pesticide program and is based on the presumption that significant lethality will not occur at concentrations below this level of toxicity. However, the slopes of dose-response curves for acute toxicity of the various pesticides must be considered in examining the validity of the assumption of negligible lethality at environmental concentrations less than or equal to 20% of the LD50. Examination of slopes for acute toxicity has shown that the criterion of 20% LD50 may be insufficiently protective for some substances,

while for others it is a worst case value and may be overly conservative. Thus, slope values of LD50 are just as important as the point estimates of lethality.

1.1.5.3 Monte Carlo and Other Probabilistic Assessment Techniques

In 1996, the U.S. EPA's Scientific Advisory Panel recommended a number of improvements in the risk assessment of pesticides, including the use of probabilistic methods. In addition, on May 15, 1997, the deputy administrator of the U.S. EPA signed a Policy for Use of Probabilistic Analysis in Risk Assessment, stating that probabilistic techniques would be used in determining ecological risk and would integrate both stressor and dose-response assessments. Such probabilistic analysis techniques are to be part of a tiered approach to risk assessment. This approach would progress from the use of simpler techniques such as quotient methods to compare point estimates of toxic effects with expected environmental exposure, to probabilistic methods that involve integration of effects and exposure distributions. Preliminary risk assessment methods using quotients are extremely useful as a screening tool to identify pesticides that may be safely used in the environment under conditions that are efficacious for their intended purpose. However, for pesticides that appear to pose significant risk, the application of Monte Carlo and other probabilistic techniques allows the analyst to account for the relationship between stressor and dose-response variables and express this relationship as likelihood of damage. Probabilistic techniques also provide a framework for expression of variability and uncertainty in risk assessments; in this way, sensitivity analyses can be performed to determine the relationship of exposure assumptions and mitigation options to risk.

The Ecological Committee on the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Risk Assessment Methods (ECOFRAM) is a peer involvement workgroup with a mission to develop probabilistic methods for pesticide risk assessment. Assessment endpoints, which are meaningful and attainable, are characterized. ECOFRAM has defined a progression of methods for risk assessment from quotients of toxicity to exposure, involving point estimates to probabilistic determinations. Initially, toxic effects are described in terms of the dose-response characteristics of a pesticide in a single test species. The slope of the dose-response curve accounts for the variance of mortality in that particular species. Retrospective analysis of toxicity information in birds and mammals has given rise to models and uncertainty factors which can be used to identify other uncertainty factors to allow for the increased sensitivity of other species (Luttik and Aldenberg, 1997; Sheehan et al. 1995). As data become available for additional species, the uncertainty factor is reduced.

Pesticide exposure assessments are based on an array of laboratory and field studies of environmental fate, which contain details regarding agricultural application rates and frequency of use. Modeling can be used to predict the range of environmental exposure levels. Monte Carlo simulation techniques are then used to integrate the dose response and exposure information. The results of risk assessment can be expressed as a probability of mortality to terrestrial nontarget populations. An estimation of the proportion of the population with at least a 90%, 75%, or 50% likelihood of dying as a result of pesticide exposure can be determined. The degree to which the distribution is sensitive to various parameters in the risk assessment model can also be examined. This aspect allows the effect of mitigation to be evaluated.

As environmental fate prediction is refined, increasing weight is given to the initial model for characterizing toxic effects of the substance to nontarget species. ECOFRAM suggests establishing additional test concentrations near the lethal threshold in acute toxicity tests to reduce variability and improve performance characteristics. In addition, to reduce the uncertainty associated with interspecies extrapolation, additional species should be tested for lethality. Approximate lethal dose methods, such as the UDP, are under consideration for this purpose. When acute toxicity studies in rats indicate that a substance poses significant risk to terrestrial mammals, an additional acute toxicity test may be required in an appropriate species of naturally occurring terrestrial populations. Similar recommendations were

made for interspecies extrapolation in avian species as part of a SETAC (Society of Environmental Toxicology and Analytical Chemistry)-OECD conference in 1994.

1.1.5.4 Endangered Species

Assessment of the potential risks of pesticides to endangered species requires that the probability of the loss of an individual be carefully assessed. An U.S. EPA agency team systematically assesses site-specific risk to endangered species using acute toxicity results. Not only is the LD50 value used, but the slope of the dose-response curve is also taken into consideration. The slope value will help to ensure that the possibility of adverse effects is carefully considered, rather than rely on a regulatory trigger based on a fixed fraction of the LD50 value. As noted above, this consideration allows the validity of assumptions of negligible risk to be tested more precisely.

1.1.6 Revised UDP

1.1.6.1 Dose Progression Factor

The current OECD UDP test guideline calls for sequential dosing with a dose progression factor of 1.3. Simulations with this progression factor clearly demonstrate that if the starting dose level is not close to the actual LD50 value for a test substance, many additional animals (as many as 30) might be needed before an adequate estimate of the LD50 is obtained. In addition, a significant bias toward the starting dose will be introduced in the results. Inclusion of a dose range-finding study was considered in order to determine the best initial dose. However, the sequential nature of dose progression in the test design of the UDP provides results that lead to centering the test doses around the LD50. Therefore, incorporation of several aspects of range-finding into the basic test was achieved by adjusting the dose spacing.

The use of simulations resulted in optimization of the test performance and increases in its applicability, by adjusting the size of the dose progression factor to 0.5 log dose (or 3.2 dose). The test should perform well with this spacing for most situations (i.e., where the slope is equal to or greater than 3.5) and will result in a more efficient use of animals.

1.1.6.2 Stopping Rule

In simulations, the number of animals needed was found to be dependent on the slope. However, in many cases, the slope is not known prior to testing and the results of the test fail to provide confidence intervals. To allow the UDP to be applied to a wide variety of test substances with reasonable reliability, the test utilizes a flexible stopping rule with criteria based on an index related to the statistical error. For test substances with higher slopes, the stopping rule will be satisfied with four animals after the first reversal. Additional animals might be needed for test substances with slopes below 4.

1.1.6.3 Limit Test

A sequential Limit Test has been designed which improves reliability of correct classification when compared to batch testing. The revised test guideline calls for attainment of three survivals or three deaths following testing at the limit dose level. In many cases, the test will be complete with three animals, although four or five animals may be needed in some cases.

1.1.6.4 Supplemental Test

A multi-sequence test has been developed as an option for determination of slope and confidence intervals. The option included in the revised guideline calls for use of multiple independent test sequences. To allow for a wide range of slope values from steep to shallow, combinations of dose progression factors can be used. To conserve animal usage, dosing for each sequence stops after reversal of outcome. Testing can be tiered in that results from the basic test can be combined with the outcome of optional testing for probit calculation of the slope and confidence intervals.

1.1.6.5 Use of a Single Sex

As agreed upon at the OECD's March 1999 Expert meeting the revised UDP uses a single sex, typically females. Female rats have a lower relative detoxification capacity for most substances, as measured by specific activity of phase I and II enzymes. Therefore, for test substances that are direct acting in their toxic mechanism, females would generally be more sensitive. If metabolic activation is required for a substance's toxicity, consideration must be given as to whether the preferred sex for testing is the male. In addition to consideration of metabolic activation and detoxification, all other information should be evaluated. Information on substance analogues or the results of testing for other toxicological endpoints of the substance itself can also indicate potential gender differences. If the investigator has reason to believe that males may be more sensitive than females, then males may be used for testing.

1.2 The Scientific Basis of the Revised UDP

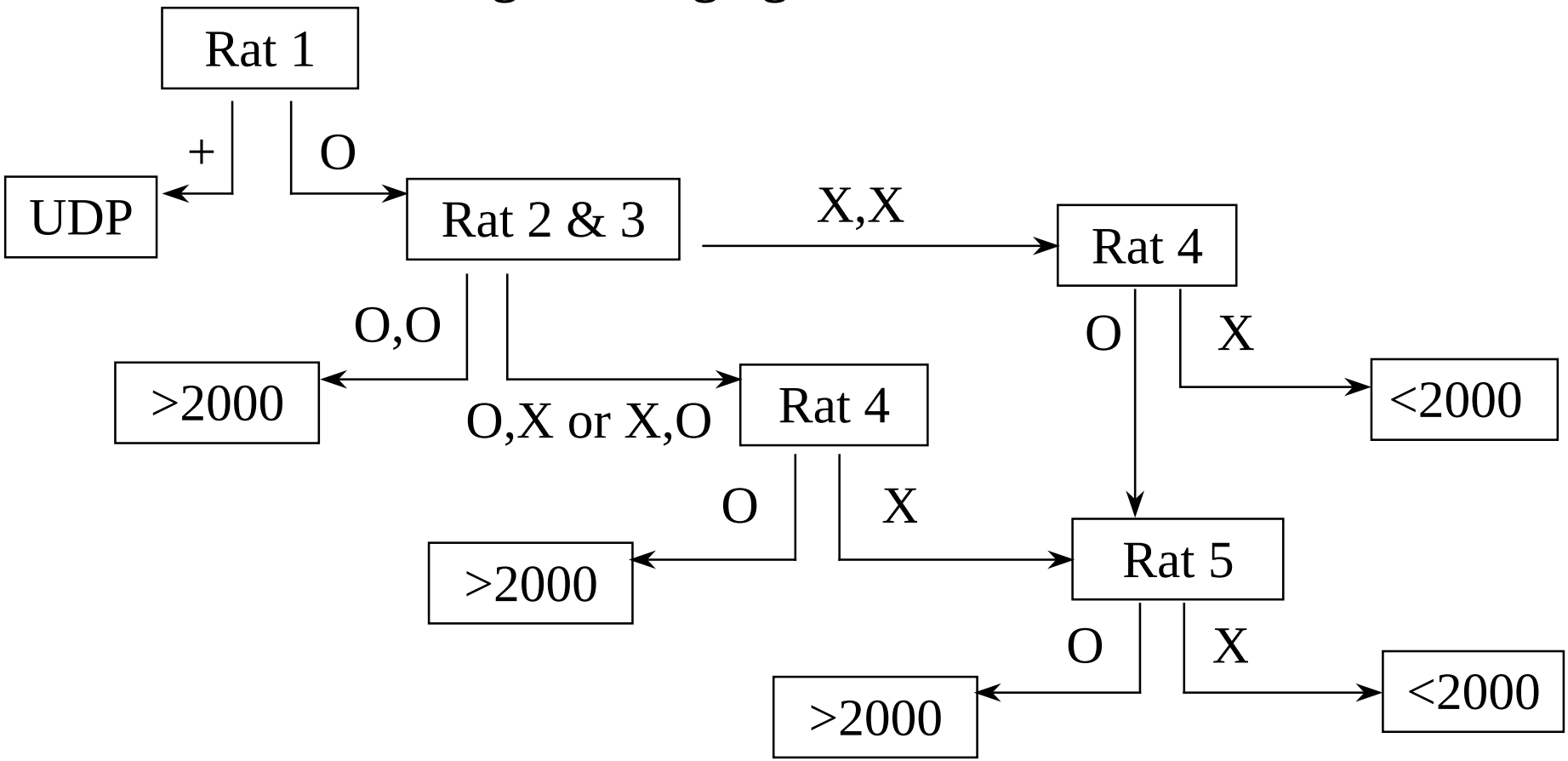
It is generally accepted that the acute oral toxicity in rats and other laboratory species can serve as an indicator of the potential for acute oral toxicity in humans. Animal studies are never perfect in their prediction of human effects; the best data for effects in humans are human data. An analysis of the historical database has demonstrated that the ranking of the LD50 values is similar between laboratory species and humans. Substances that are not toxic in the rat are often not toxic in humans and substances that are highly toxic in the rat are often highly toxic in humans. Since human testing for acute lethality is unethical and illegal, animal bioassays have provided data that are reasonable approximations of the effects in humans. The revised UDP method permits estimation of an LD50 with a confidence interval and the results allow a substance to be ranked and classified according to the OECD Globally Harmonised System for the classification of substances that cause acute toxicity.

The primary 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 level a step below the level of the best estimate of the LD50. If the animal survives, the dose level for the next animal is increased to a default factor of 3.2 times the original dose level; if it dies, the dose level for the next animal is decreased by a similar dose progression 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--a decision which is based on the 48-hour survival pattern of all the animals up to that time. 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. Dosing is stopped when one of these criteria is satisfied, at which time an estimate of the LD50 and a confidence interval are calculated for the test based on the status of all the animals at termination. For most applications, testing will be completed with only 4 animals after initial reversal in animal outcome. The LD50 is calculated using the method of maximum likelihood.

The Limit Test is a sequential test that uses a maximum of five animals. A test dose of up to 2000 or, exceptionally, 5000 mg/kg, may be used. The selection of a sequential test plan increases the statistical power and also has been made to intentionally bias the procedure toward rejection of the limit test for test

substances 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. Figure 1-1 shows a flowchart schematic for the UDP Limit Test procedure.

Figure 1-1. Flowchart Schematic for the UDP Limit Test Procedure, using 2000 mg/kg as the Limit Dose



Note: O equates to non-toxic and X equates to toxic

1.3 Intended Regulatory Uses of the Revised UDP

The regulatory basis for the Revised UDP is the need to identify the toxic effects of a given test substance as part of a safety evaluation for potentially exposed humans. Acute toxicity testing provides information on the health hazards likely to arise from short-term exposure and is typically an initial step in the evaluation of the toxic characteristics of a chemical substance. Data from acute studies may serve many different roles, such as to:

- provide a basis for hazard classification and labeling
- establish dosing levels for repeated-dose toxicity studies
- generate information on affected organs
- give clues as to the mode of toxic action
- aid in the diagnosis and treatment of toxic reactions
- provide information for comparison of toxicity and dose response among members of chemical classes
- help standardize biological products
- serve as a standard for evaluating alternatives to the animal test
- help judge the consequences of exposures in the workplace, at home, and on accidental release

The Revised UDP will replace the current regulations on acute oral toxicity testing for the CPSC, the U.S. EPA, and the U.S. Department of Transportation (DOT). The Revised UDP will specifically provide the following:

1. Point Estimate of Lethality for Classification:
   - classification of pure substances - CPSC, DOT, Occupational Safety and Health Administration (OSHA)
   - classification of mixtures - CPSC, DOT, OSHA
   - classification of pesticide active ingredients and formulations - U.S. EPA
   - characterization of inerts in pesticide formulations – U.S. EPA

2. Range Estimate of Lethality for Classification:
   - classification of pure substances - CPSC, DOT, OSHA
   - classification of pesticide formulations – U.S. EPA

3. Risk Assessment (Slope, Confidence Intervals, Dose-Effect)
   - human health assessment, pure substances and mixtures - CPSC, OSHA; and pesticides – U.S. EPA
   - environmental assessment of pesticides – U.S. EPA

4. Limit Dose at 5000 mg/kg:
   - Pesticides, safer chemical policy/incentives, biological agents – U.S. EPA
   - consumer products - CPSC

Because the Revised UDP provides an estimate of the slope of the dose-response curve and the confidence interval for the LD50, the data can also be used for risk assessment purposes and probabilistic modeling.

1.4 Currently Accepted Acute Oral Toxicity Test Methods

Should the Revised UDP be adopted by the OECD, it is expected that U.S. Federal agencies requiring

acute toxicity data as generated by OECD TG 401 will accept the UDP as the alternative acute oral toxicity test. Guidelines and regulations for acute oral toxicity are shown in **Table 1-1**. The current guidelines of U.S. Federal agencies for acute oral testing are:

1. Under the Federal Hazardous Substances Act, the CPSC requires testing of groups of 10 rats weighing between 200 and 300 g at doses between 50 and 5000 mg/kg followed by a 14-day observation period to obtain an LD50 (16 CFR 1500; original BRD **Appendix E**, final report **Appendix Q-1**). OECD TG 401 is an accepted test method. For the Limit Test, a group of 10 rats is dosed at 5000 mg/kg and observed for 14 days.
2. Under FIFRA, the U.S. EPA requires the testing of rats weighing between 200 and 300 g at doses between 5 and 5000 mg/kg followed by a 14-day observation period (40 CFR 152; original BRD **Appendix E**, final report **Appendix Q-3**). OECD TG 401 and TG 425 are accepted test methods.
3. Under FIFRA, the U.S. EPA requires the identification of the range of the acute oral LD50s by testing rats weighing between 200 and 300 g followed by a 14-day observation period (40 CFR 156; original BRD **Appendix E**, final report **Appendix Q-4**). OECD TG 401, TG 420, TG 423, and TG 425 are accepted test methods.
4. Under FIFRA, the U.S. EPA requires acute oral testing of chemicals and products which may become a residue in food and nonfood crops (40 CFR 158; original BRD **Appendix E**, final report **Appendix Q-5**). OECD TG 401 and TG 425 are accepted test methods.
5. Under the Toxic Substances Control Act (TSCA), the U.S. EPA requires acute oral toxicity data for chemicals proposed for a significant new use (40 CFR 721; original BRD **Appendix E**, final report **Appendix Q-6**). OECD TG 401 and TG 425 are accepted test methods.
6. The U.S. DOT and its 11 administrations require the identification of the range of the acute oral LD50s by testing in young adult rats (49 CFR 173; original BRD **Appendix E**, final report **Appendix Q-7**). OECD TG 401, TG 420, TG 423, and TG 425 are accepted test methods.

For the U.S. EPA OPP, the LD50 for a test substance may be obtained using several methods including, (1) OECD TG 401 in which three groups of five female rats, 8 to 12 weeks of age, receive a single oral dose of the test substance and are observed for 14 days with a single confirming dose given to five male rats; (2) a conventional LD50 test in which several groups of five male and five female rats are given a single oral dose of the test substance and are observed for 14 days, with the selected dose levels based on a range-finding study, and (3) the UDP method can be used, but requires the submission of an acceptable protocol (e.g., OECD TG 425). In addition, a Limit Test may be conducted for a group of five male and five female rats given a single oral dose of 2000 or 5000 mg/kg and observed for 14 days.

Table 1-1 Guidelines and Regulations for Acute Oral Toxicity

AGENCY OR ORGANIZATION	GUIDELINES AND REGULATIONS[1]	COMMENTS
Consumer Product Safety Commission (CPSC)	**16CFR1500** Hazardous Substances and Articles: Administration and Enforcement	The CPSC, as mandated under the Federal Hazardous Substances Control Act, requires acute oral toxicity and other testing be conducted on chemicals in commerce. The purpose is to provide adequate labeling and warning to consumers of goods that are hazardous via oral, dermal, or inhalation during purposeful or accidental exposure.
	§1500.3 Definitions	A single oral dose in rats followed by a 14-day observation period, for classification purposes.
U.S. Department of Transportation (U.S. DOT)	**49CFR173** Shippers – General Requirements for Shipments and Packaging	The DOT, in compliance with Hazardous Materials Regulations, outlines the requirements to be observed in preparing hazardous materials for shipment by air, highway, rail, or water, or any combination thereof. These regulations are based on the Recommendations of the United Nations Committee of Experts on the Transport of Dangerous Goods, the International Civil Aviation Organization, and the International Maritime Organization.
	§173.132 Definitions **§173.133** Assignment of packing group and hazardous zones for Division 6.1 materials	Classification based on LD50 for packing requirements.
U.S. Environmental Protection Agency (EPA) Office of Pesticide Programs (OPP)	**40CFR152** Pesticide Registration and Classification Procedures	The U. S. EPA is required under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) to register all pesticides available for use in the U.S. This section sets forth the procedures, requirements, and criteria for registration and reregistration of pesticide products, and regulatory activities affecting registration. Testing must be in compliance with Good Laboratory Practices (GLPs) (40 CFR Part 792).
	§152.3 Definitions	A statistical-derived estimate of the single oral dose level of a substance causing 50% mortality to the test population under specified conditions.
	§156.10 Labeling requirements for Pesticides and Devices	The U. S. EPA is required under FIFRA to adequately label all pesticide products for use in the U.S. Such labeling is primarily for worker protection and must include information on toxicity, symptoms, treatment, and recommended personal protective equipment. Testing must be in compliance with GLPs (40 CFR Part 792). Classification based on the LD50 for labeling requirements.
	§158.20 Data Requirements for Registration	This section specifies the types and amounts of data and information required by the Agency to make informed decisions on the risks and benefits of various pesticide products. Testing must be in compliance with GLPs (40 CFR Part 792). An acute oral LD50 is part of the minimum data package for registration.
	§158.70 Acceptable protocols	OECD protocols can be used to develop data necessary to data requirements.

U.S. Environmental Protection Agency (U.S. EPA) Office of Pesticide Programs (OPP)	**40CFR721** Significant new uses of chemical substances	The U. S. EPA requires vendors under the Toxic Substances Control Act (TSCA) to conduct acute oral toxicity studies according to harmonized test guidelines (OECD TG 401). A safety evaluation must be conducted for each proposed new use of a chemical substance. Testing must be in compliance with GLPs (40 CFR Part 792).
U.S. EPA, Office of Pollution Prevention and Toxic Substances (OPPTS)	**OPPTS 870.1100** Acute Oral Toxicity	EPA Health Effects Test Guidelines http://www.epa.gov/docs/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Drafts/

[1] Unless otherwise specified in the comments column, guidelines may be accessed via the U.S. Government Printing Office (GPO) Code of Regulations database http://www.access.gpo.gov/nara/cfr/cfr-table-search.html.

1.5 Intended Range of Substances Amenable to Testing Using the Revised UDP

Because the method of dosing (i.e., oral gavage) is the same for OECD TG 401 and the Revised UDP, any class of substances and products that can or have been tested using TG 401 can be tested using the Revised UDP. The test is designed for substances that can be administered neat (i.e., without dilution) or in a solvent. The test is not restricted to water-soluble substances. Any solvent or vehicle can be used, but the solvent or vehicle must not add to or mask the toxicity of the test substance.

2.0 Proposed Protocol for the Revised UDP

2.1 Detailed Protocol and Rationale

OECD adopted the UDP as TG 425 in October 1998 (original BRD **Appendix A,** final report **Appendix H)**. The UDP Primary test has now been revised by changing the default starting dose level, the dose-spacing factor, the time period before the dosing of the next animal, and the stopping criteria. The UDP Limit Test was changed to utilize females only and to allow, for specific regulatory purposes, a limit dose level of 5000 mg/kg. In addition, an UDP Supplemental Test has been added to provide the estimation of the slope of the dose-response curve and the 95% confidence interval of the LD50. The Revised UDP guideline has been prepared using OECD test guideline format and is entitled, "Acute Oral Toxicity: Modified Up-and-Down Procedure (Revised UDP)" (see U.S. EPA Document 1B **–** original BRD **Appendix C**, final report **Appendix G)**. A description of the Revised UDP follows; exact wording from the UDP guideline (version 425N) is set in quotation marks.

2.1.1 Materials, Equipment, and Supplies

2.1.1.1 Selection of animal species

"The preferred rodent species is the rat although other rodent species may be used. In the normal procedure, female rats are used because literature surveys of conventional LD50 tests show that, although there is little difference of sensitivity between sexes, in those cases where differences were observed, females were in general more sensitive. When there is adequate information to infer that males are more sensitive, they should replace females in the test" (see paragraph 12, Revised UDP, U.S. EPA Document 1B **-** original BRD **Appendix C**, final report **Appendix G**).

This section has not been altered from that provided in the original UDP.

"Healthy young adult animals should be employed. Littermates should be randomly assigned to treatment levels. The females should be nulliparous and non-pregnant. At the commencement of the study, the weight variation of the animals should be minimal and not exceed ±20% of the mean weight for each sex. The test animals should be characterized as to species, strain, source, sex, weight and/or age" (see paragraph 13, Revised UDP, U.S. EPA Document 1B – original BRD **Appendix C**, final report **Appendix G**).

Because the UDP requires at least 48 hours between the sequential dosing of animals, the ±20% variation rule for body weight may too restrictive. Utilizing animals from the same shipment in a randomized manner in which dosing may occur over a two to three week period may result in many animals exceeding this specified weight range, leading to increased animal use and associated costs.

2.1.1.2 Housing and feeding conditions

"The temperature in the experimental animal room should be 22%C (±3%C). Although the relative humidity should be at least 30% and preferably not exceed 60% other than during room cleaning, the aim should be 50-60%. Lighting should be artificial, the sequence being 12 hours light and 12 hours dark. The animals are housed individually. Unlimited supply of conventional rodent laboratory diets and drinking water should be provided" (see paragraph 14, Revised UDP, U.S. EPA Document 1B - original BRD **Appendix C**, final report **Appendix G**).

This section has not been altered from that provided in the original UDP.

2.1.1.3 Preparation of animals

"The animals are uniquely identified and kept in their cages for at least five days prior to dosing for acclimatization to the laboratory conditions. During acclimatization the animals should be observed for ill health. Animals demonstrating signs of spontaneous disease or abnormality prior to the start of the study are eliminated from the study" (see paragraph 15, Revised UDP, U.S. EPA Document 1B - original BRD **Appendix C**, final report **Appendix G**).

This section has not been altered from that provided in the original UDP.

2.1.1.4 Preparation of doses

"When necessary, the test substance is dissolved or suspended in a suitable vehicle. It is recommended that, whenever possible, the use of an aqueous solution or suspension be considered first, followed by consideration of a solution or emulsion in oil (e.g., corn oil) and then by possible solution in other vehicles. For vehicles other than water, the toxicity of the vehicle must be known. In rodents, the volume should not normally exceed 1 mL/100 g body weight; however, in the case of aqueous solutions 2 mL/100 g body weight can be considered." (see paragraph 16, Revised UDP, U.S. EPA Document 1B - original BRD **Appendix C**, final report **Appendix G**).

This section has not been altered from that provided in the original UDP.

2.1.2 Procedure

2.1.2.1 Primary testing using a single-sequence of dosing

"For selecting the starting dose, all available information should be used, including information on structure-activity relationships. When the information suggests that mortality is unlikely, a limit test should be conducted. When there is no information on the substance to be tested, it is recommended that the starting dose of 175 mg/kg body weight be used. This dose serves to reduce the level of pain and suffering by starting at a dose level which in most cases will be sublethal. In addition, this dose reduces the chance that hazard of the chemical will be underestimated" (see paragraph 17, Revised UDP, U.S. EPA Document 1B - original BRD **Appendix C**, final report **Appendix G**).

"For each run, single animals are dosed in sequence usually at 48-hour intervals. However, the time intervals between dosing should not be fixed rigidly and may be adjusted as appropriate (e.g., in case of delayed mortality). The first animal is dosed a step below the toxicologist's best estimate of the LD50. If no estimate of the chemical's lethality is available, dosing should be initiated at 175 mg/kg. 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. Animals killed for humane reasons are considered in the same way as animals that died on test. Dosing should not normally exceed 2000 mg/kg body weight or 5000 mg/kg body weight as justified by specific regulatory needs" (see paragraph 18, Revised UDP, U.S. EPA Document 1B - original BRD **Appendix C**, final report **Appendix G**).

Prior to conducting the study, the testing laboratory should consider all available information on the test substance. Such information will 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; toxicological data on structurally related substances or similar mixtures; and the anticipated use(s) of the substance. This information is useful to determine the relevance of the test for the protection of human health and the environment, and will help in the selection of an appropriate starting dose.

The UDP suggested a dosing sequence of 24 hours. Since some animals die between 24 and 48 hours post-dosing and because fasting of the next animal to be dosed typically does not start until at least 24 hours after the treatment of the preceding animal, the dosing sequence in the revised UDP is at least 48 hours.

"Moribund state is characterized by symptoms such as shallow, labored or irregular respiration, muscular weakness or tremors, absence of voluntary response to external stimuli, cyanosis, and coma. Criteria for making the decision to humanely kill moribund and severely suffering animals are the subject of the separate OECD *Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation"* (see paragraph 19, Revised UDP, U.S. EPA Document 1B - original BRD **Appendix C**, final report **Appendix G**). The Guidance Document was provided the original BRD as **Appendix B**, but is not appended to this final report.

The Revised UDP emphasizes careful cageside and in-hand observations as described in the Guidance Document.

2.1.2.2 Dose-Spacing Factor and Stopping Rules

"The dose for each successive animal is adjusted up or down, depending on the outcome of the previous animal. At the outset, if feasible, a slope of the dose response should also be estimated based on all information available to the toxicologist including structure activity relationships. The dose progression factor should be chosen to be the antilog of 1/(the estimated slope of the dose-response curve). When there is no information on the substance to be tested, a dose progression factor of 3.2 is used. Dosing continues depending on the outcomes of all the animals up to that time. In any event, if 15 animals have been tested, testing stops. Prior to that, the test is stopped based on the outcome pattern if:

1) the upper testing bound is reached and 3 consecutive animals survive at that bound or if the lower bound is reached and 3 consecutive animals die at that bound, or
2) the next animal to be tested would be the 7th and each surviving animal to this point has been followed by a death and vice versa (i.e., 5 reversals occur in 6 animals started), otherwise;
3) evaluation whether testing stops or continues is based on whether a certain stopping criterion is met: Starting following the fourth animal after the first reversal (which may be as early as the decision about the seventh animal), three measures of test progress are compared via two ratios. If the first measure is at least two-and-one-half times both of the other measures (i.e., both ratios are 2.5), testing is stopped.

   For a wide variety of combinations of LD50 and slopes as low as 2.5, the stopping rule will be satisfied with four to six additional animals, with fortuitously well-placed tests using even fewer. However, for chemicals with shallow dose-response slope (large variance), more animals may be needed. If animal tolerances to the chemical are expected to be highly variable (i.e., slopes are expected to be less than 3), consideration should be given to increasing the dose progression factor beyond the default 0.5 log dose (i.e., 3.2 progression factor) prior to starting the test."

   When the stopping criteria have been attained after the initial reversal, the LD50 should be calculated using the method described in" Section 2.1.7.3 (see paragraph 20 and 21, Revised UDP, U.S. EPA Document 1B - original BRD **Appendix C**, final report **Appendix G**).

In the current UDP, the dose-spacing factor was 1.3. This factor has been changed to 3.2 in the Revised UDP because:

1. if the starting dose level is far from the LD50, a dose-spacing factor of 1.3 may use many animals to reach the LD50; and
2. if the dose-response curve is very shallow (2.5 or less), a factor of 1.3 leads to a significant possibility of bias toward the starting dose level.

For example, if the LD50 is 1878 mg/kg and the starting dose level is 175 mg/kg, it would require 12 animals to approach the LD50. A spacing factor of 3.2 requires the use of only three animals. If the slope is shallow and the starting dose level is far from the LD50, it is likely that there will be a reversal of outcome far from the LD50. Since the current UDP stops with four animals after the first reversal, the test often does not reach the LD50 prior to meeting stopping criteria. A complete description of the development of the stopping criteria is given in U.S. EPA Document 5 (original BRD **Appendix C**, final report **Appendix K**).

2.1.3 The Supplemental Test: Estimate of an LD50 and Slope of the Dose-Response Curve

"Following the primary test, a supplemental test to estimate the slope of the dose-response curve can be implemented when necessary. This procedure uses multiple testing sequences similar to the primary test, with the exception that the sequences are intentionally begun well below the LD50 estimate from the primary test. These test sequences should be started at doses at least 10 times less than the LD50 estimate from the primary test and not more than 32 times less. Testing continues in each sequence until the first animal dies. Doses within each sequence are increased by the standard 3.2 factor. The starting dose level for each test sequence should be staggered, as described in Appendix II, paragraph 6. Upon completion of up to six of these supplemental test sequences, a standard probit analysis should be run on the entire collection of data, including the outcomes of the primary test. Good judgment will be required in cases where the primary test yields estimates of LD50 that are too close to the lower limit of doses tested. When this occurs, testing may be required to begin well above the LD50, where deaths are likely, and each sequence will terminate with the first survivor. If slope may be highly variable, an alternate

procedure, using varying dose progression sizes, may be appropriate" (see paragraph 22, Revised UDP, U.S. EPA Document 1B - original BRD **Appendix C**, final report **Appendix G**).

A complete description of the development of the Supplemental Test is given in U.S. EPA Document 8 (original BRD **Appendix C**, final report **Appendix N**).

2.1.4 The Limit Test

"Dosing should not normally exceed 2000 mg/kg body weight. However, when justified by specific regulatory needs, testing up to 5000 mg/kg body weight may be considered. One animal is dosed at the upper limit dose; if it survives, two more animals are dosed sequentially at the limit dose; if both animals survive, the test is stopped. If one or both of these two animals die, two animals are dosed sequentially at the limit dose until a total of three survivals or three deaths occur. If three animals survive, the LD50 is estimated to be above the limit dose. If three animals die, the LD50 is estimated to be at or below the limit dose. If the first animal dies, a primary test should be run to determine the LD50." A flow chart delineating the procedures for the Revised UDP Limit Test is shown in **Table 2-1**.

"As with any limit test protocol, the probability of correctly classifying a compound will decrease as the actual LD50 approaches the limit dose. The selection of a sequential test plan increases the statistical power and also has been made to intentionally bias the procedure toward rejection of the limit test for compounds with LD50 values near the limit dose (i.e., to err on the side of safety)" (see paragraph 23, Revised UDP, U.S. EPA Document 1B; original BRD **Appendix C**, final report **Appendix G**).

In the Revised UDP, the test stops when testing is complete in females; whereas, in the current UDP, three males are tested following testing in females. A complete description of the rationale for the Limit Test is given in U.S. EPA Document 7 (original BRD **Appendix C**, final report **Appendix M**).

Table 2-1 Flow Chart for the Revised UDP Limit Test

1.	Test one animal (first animal)	- -	if it survives, then test two additional animals if it dies, then conduct the Primary Test
2.	Test two animals (second and third animals)	-	if both survive, then the test is complete if one or both die, then test two additional animals sequentially
3.	Test two animals sequentially (fourth and fifth animals)	-	stop the test as soon as three animals have survived or died. If three animals have died, then conduct the Primary Test

2.1.5 Dosing Procedures

2.1.5.1 Administration of doses

"The test substance is administered in a single dose to the animals by gavage using a stomach tube or a suitable intubation cannula. 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/100 g body weight; however, in the case of aqueous solutions 2 ml/100 g body weight can be considered. When a vehicle other than water is used, variability in test volume should be minimized by adjusting the concentration to ensure a constant volume at all dose levels. If administration in a single dose is not possible, the dose may be given in smaller fractions over a period not exceeding 24 hours.

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" (see paragraphs 24 and 25, Revised UDP, U.S. EPA Document 1B - original BRD **Appendix C**, final report **Appendix G**).

This section has not been altered from that provided in the original UDP.

2.1.6 Endpoints Recorded

2.1.6.1 Observations

"After dosing, animals are observed individually at least once during the first 30 minutes, periodically during the first 24 hours, with special attention given during the first 4 hours, and at least once daily thereafter. The animals should normally be observed for 14 days, except where animals 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. The length of the observation period should be determined by the toxic reactions, time of onset, and length of recovery period, and may thus be extended when considered necessary. The times at which signs of toxicity appear and disappear are important, especially if there is a tendency for toxic signs to be delayed. All observations are systematically recorded with individual records being maintained for each animal. Toxicology texts should be consulted for information on the types of clinical signs that might be observed" (see paragraph 26, Revised UDP, U.S. EPA Document 1B - original BRD **Appendix C**, final report **Appendix G**).

In the revised UDP, more emphasis is placed on humane endpoints and clinical signs. Examples of clinical signs were provided in the original BRD in **Appendix B**; this appendix is not included in this final report.

"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. 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. 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" (see paragraph 27, Revised UDP, U.S. EPA Document 1B - original BRD **Appendix C**, final report **Appendix G**).

More emphasis is placed on humane endpoints and clinical signs in the Revised UDP. Humane treatment of animals was described in the original BRD in **Appendix B**; this appendix is not included in this final report.

2.1.6.2 Body weight

"Individual weights of animals should be determined shortly before the test substance is administered, at least weekly thereafter, at the time of death or at day 14 in the case of survival. Weight changes should

be calculated and recorded" (see paragraph 28, Revised UDP, U.S. EPA Document 1B - original BRD **Appendix C**, final report **Appendix G**).

This section has not been altered from that provided in the original UDP.

2.1.6.3 Pathology

"All animals, including those which die during the test or are killed for animal welfare reasons during the test and those that survive at day 14, are subjected to gross necropsy. The necropsy should entail a macroscopic inspection of the visceral organs. As deemed appropriate, microscopic analysis of target organs and clinical chemistry may be included to gain further information on the nature of the toxicity of the test material" (see paragraph 29, Revised UDP, U.S. EPA Document 1B - original BRD **Appendix C**, final report **Appendix G**).

This section has not been altered from that provided in the original UDP.

2.1.7 Data and Reporting

2.1.7.1 Data

"Individual animal data should be provided. Additionally, all data should be summarized in tabular form, showing the following for each test concentration: the number of animals used; the number of animals displaying signs of toxicity; the number of animals found dead or killed for humane reasons; time of death for each animal; 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" (see paragraph 30, Revised UDP, U.S. EPA Document 1B - original BRD **Appendix C**, final report **Appendix G**).

This section has not been altered from that provided in the original UDP.

2.1.7.2 Data Storage

Original data are collected and maintained in study books according to Agency-accepted Good Laboratory Practices (GLPs). Data are then entered into computerized spreadsheets for manipulation and analysis.

2.1.7.3 Calculation of LD50 for the Primary Test

"The LD50 is calculated using the maximum likelihood method, other than in exceptional cases given below. 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 (1991a), 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 is the total number of animals tested.

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

where

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

When identifying the maximum of the likelihood *L* to get an estimate of the true LD50, *mu* is set = log LD50 and automated calculations solve for it.

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

(a) If testing stopped based on criterion (1) (i.e., a boundary dose was tested repeatedly; see Section 2.1.2.2), or if the upper bound dose ended testing, then the LD50 is reported to be above the upper bound; if the lower bound dose ended testing then the LD50 is reported to be below the lower bound dose. Classification is completed on this basis.

(b) If all the dead animals have higher doses than all the live animals, or vice versa, 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 (2) (i.e., 5 reversals occur in 6 animals started; see Section 2.1.2.2) 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 there is ever cause to repeat the test, 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.

Maximum likelihood calculation can be performed using either SAS (e.g., PROC NLIN) or BMDP (e.g., program AR) computer program packages as described (SAS, 1990; BMDP, 1990). Other computer programs may also be used. Typical instructions for these packages are given in appendices to the American Society for Testing and Materials (ASTM) Standard E 1163-87. The *sigma* used in the BASIC program will need to be edited to reflect the changes in this version of the OECD 425 Guideline. The program's output is an estimate of log(LD50) and its standard error.

The stopping criterion (3) (i.e., is based on three measures of test progress that are of the form of the likelihood (see Section 2.1.2.2) with different values for *mu*, and comparisons are made after each animal tested after the sixth that does not already satisfy criterion (1) or (2). The equations for criterion (3) are provided in Appendix III. 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 III. If the criterion is met, testing stops and the LD50 can be calculated by the maximum likelihood method" (see paragraph 31 to 33, Revised UDP, U.S. EPA Document 1B - original BRD **Appendix C**, final report **Appendix G**).

After the sixth animal is dosed, the stopping rule is checked after each additional animal is tested. When the stopping rule is satisfied, the LD50 is calculated.

2.1.7.4 Calculation of LD50 and Slope Using Supplemental Procedure

"A Supplemental Procedure is based on running three independent replicates of the Up-and-Down Procedure. Each replicate starts at least one log, but not greater than 1.5 log, below the estimated LD50. Each run stops when the first animal dies. All data from these runs and the original Up-and-Down run are combined and an LD50 and slope are calculated using a standard probit method" (see paragraph 34, Revised UDP, U.S. EPA Document 1B - original BRD **Appendix C**, final report **Appendix G**).

No statistical procedures are required for the Limit Test.

2.1.8 Report

"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;
- slope of the dose-response curve (when determined);
- LD50 data;
- statistical treatment of results (description of computer routine used and spreadsheet tabulation of calculations)

Discussion and interpretation of results

Conclusions:

(see paragraph 35, Revised UDP, U.S. EPA Document 1B - original BRD **Appendix C**, final report **Appendix G**).

This section has not been altered from that provided in the original UDP.

2.1.9 Equipment and Training

2.1.9.1 Equipment

Equipment needed is the same as the standard equipment for any oral toxicity test, including: cages, balances, analytical equipment as necessary to confirm the identity of the test substance, possibly waterbaths or mixers to dissolve the substance, dosing syringes, gavage catheters, and necropsy equipment. The only special piece of equipment needed for this method is a standard personal computer to run a spreadsheet program and a means to run maximum likelihood estimates using SAS or a similar program. It is anticipated that stopping rule program will be made available in Excel® or some other standard format on the OECD or U.S. EPA websites or on a floppy disk. It could also be written, as described in the guideline, by the toxicologists themselves if preferred.

2.1.9.2 Training

Technicians running the Revised UDP must be trained to properly calculate, mix, and administer test substances to rats via oral gavage and trained to make and record observations in an acute toxicity study, including the gross necropsy. They should also be familiar with OECD guidelines on humane endpoints and able to make decisions on when to sacrifice a terminally ill animal.

Staff must also be able to use the computer programs. A full description of how to use the stopping rule, with examples, is included in the guideline. The use of the maximum likelihood method for calculating the LD50 is a standard statistical program and would require experience in these programs. Training may be available for those unfamiliar with this type of computer program. Dosing and observations are similar to other acute toxicity protocols. For all acute toxicity studies, technicians conducting the studies must be trained in making and recording observations correctly; this training is a very important aspect of the guideline and is often overlooked.

2.1.10 Basis for the Selection of Females

In revising TG 401 in 1987, OECD required the use of only one sex of the test species. Differences in gender sensitivity may include, but are not limited to, differences in specific enzyme systems (e.g., cytochrome P450 or conjugation pathways) and differences in absorption, distribution, and excretion (e.g., body fat content and distribution). A complete discussion of gender considerations is given in U.S. EPA Document 14 (original BRD **Appendix C**, final report **Appendix P**).

2.1.11 Confidential Information

There are no confidential data associated with the Revised UDP.

2.1.12 Decision Criteria for the Revised UDP

The decision criteria for the Revised UDP are detailed in the test guideline. Decision criteria for an adequate test and for stopping testing are proposed to be part of the computer program (see U.S. EPA Document 6 - original BRD **Appendix C**, final report **Appendix L**).

2.2 Basis for the Number of Replicate and Repeat Experiments

Historically, only a single experiment has been required to estimate the LD50 for a test substance (see OECD TG 401, TG 425, Revised UDP). The scientific basis for this requirement is unknown, but is most likely based on limiting animal use and the realization that the resulting LD50 is only a reasonable approximation. Similarly, the Limit Test is based on a single test. In contrast, the Supplemental Test in the Revised UDP, in order to calculate the slope of the dose-response curve and the corresponding confidence interval of the LD50, is based on three to four replicate tests. The justification for this number of replications is provided in U.S. EPA Document 1B (original BRD **Appendix C**, final report **Appendix G**).

2.3 Protocol Modifications as a Result of Validation Studies

The Revised UDP is a test guideline constructed and validated using computer simulations. The computer simulation studies were used to optimize the protocol as to starting dose level, dose-spacing factor, and stopping rules. The starting dose level has been changed to 175 mg/kg as part of the process to reduce animal use for test substances with a shallow slope in the dose-response curve. The dose-spacing factor was increased to 3.2 to curtail excess animal use prior to the first reversal when the starting dose level is far from the LD50. The stopping criteria allow for a more accurate estimate of the LD50 for test substances with a shallow slope and yet require only six or seven animals when the slope is steep.

3.0 Characterization of the Substances Tested

Three *in vivo* studies have been conducted using the UDP. The test substances used in each study are presented below. For the Bruce (1987) study, selection of the test substances was based on a wide variation in LD50 values (from 273 to more than 20,000 mg/kg). The rationale for selecting the five substances in the Bonnyns et al. (1988) study was that each compound affected different target organs; the published LD50 values ranged between 200 to 2000 mg/kg. In the Yam et al. (1991) study, the ten compounds were arbitrarily selected from the 20 test substances studied by van den Heuvel (1990), with consideration given to the range of LD50 values (48 to greater than 3000 mg/kg).

Table 3-1 Reference Test Substances

Bruce (1987)

Test Substance	Chemical/Product Class	CAS Number
Proprietary	Ingredient	-
Proprietary	Laundry detergent	-
Proprietary	Ingredient	-
Proprietary	Laundry detergent	-
Proprietary	Laundry detergent	-
Proprietary	Shampoo	-
Proprietary	Flavor	-
Caffeine	Stimulant	58-08-2
Potassium hydroxide	Strong base	1310-58-3
Proprietary	Dishwashing detergent	-

Bonnyns et al. (1988)

Test Substance	Chemical/Product Class	CAS Number
Barium acetate	Metal salt	543-80-6
Barbital	CNS depressant	57-44-3
Coumarin	anticoagulant drug	91-64-5
Allyl heptanoate	alkyl ester	-
Diquat	Herbicide	85-00-7

Yam et al. (1991)

Test Substance	Chemical/Product Class	CAS Number
Nicotine	plant product	54-11-5
Na pentachlorophenate	chlorinated organic salt	-
Na arsenite	metal salt	7784-46-5
p-Dichlorobenzene	chlorinated solvent	106-46-7
Fentin hydroxide	organic tin fungicide	76-87-9
Acetanilide	medicinal/intermediate	103-84-4
Tetrachlorvinphos	organophosphate pesticide	-
Piperidene	solvent	110-89-4
Mercuric chloride	metal salt	7487-94-7
4-Aminophenol	solvent	123-30-8

4.0 Reference Data Used for Performance Assessment

In LD50 studies using OECD TG 401, it was common practice to dose 50 or more animals simultaneously and evaluate lethality based on a 14-day observation period. The UDP involves the dosing of animals in a sequential manner. Sequential sampling is a novel approach to LD50 testing, although it has been used successfully in other areas. Bruce (1987) evaluated the UDP using a series of ten substances and the results were compared with LD50 values generated using TG 401. In this series, the test substances consisted primarily of surfactant-based cleaners, but also included a flavoring substance, caffeine, and potassium hydroxide. Subsequently, two other studies (Bonnyns et al., 1988; Yam et al., 1991) compared the results of the UDP with the classical LD50 test (OECD TG 401). In the Yam et al. (1991) study, the OECD TG 401 data used for comparison were taken from the van den Heuvel et al. (1990) study. In total, 25 substances were evaluated in these studies, as detailed in Lipnick et al. (1995). This number of compounds for validation studies is similar to that run for the FDP (20 compounds) (van den Heuvel et al., 1990) and the ATCM (30 compounds) (Schlede et al., 1992).

4.1 Protocol for Reference Data (OECD TG 401)

The reference data were generated using OECD TG 401. No deviations to the protocol were noted in the Bruce (1987), Bonnyns et al. (1988), or the van den Heuvel (1990) studies.

4.2 Results for OECD TG 401 Studies

Listings of the substances in the three comparison studies of the UDP are provided in **Table 4-1**. In the Bruce (1987) and the Bonnyns et al. (1988) studies, the authors simultaneously conducted acute oral testing using OECD TG 401. The Yam et al. (1991) study was part of the validation study for FDP and the OECD TG 401 data for both studies were taken from the van den Heuvel (1990) study.

Table 4-1 Results from TG 401 Studies

Test Substance	LD50 (mg/kg)
Bruce (1987)	
Ingredient	>20,000
Laundry detergent	10,110
Ingredient	>10,000
Shampoo	9,280
Dishwashing detergent	5,560
Laundry detergent	4,040
Laundry detergent	3,510
Flavor	3,490
Caffeine	344
Potassium hydroxide	273
Bonnyns et al. (1988)	
Diquat	1,036
Allyl heptanoate	991
Barium acetate	571
Coumarine	470
Barbital	404
Yam et al. (1991)	
4-Aminophenol	>3,000
p-Dichlorobenzene	>2,000
Tetrachlorvinphos	>2,000
Acetanilide	1,893
Piperidene	488
Na pentachlorophenate	309
Mercuric chloride	160
Fentin hydroxide	119
Nicotine	71
Na arsenite	48

4.3 Original Data Sheets

Proctor and Gamble Company provided original datasheets for portions of the Bruce (1987) and the Yam et al. (1991) studies. Additional original datasheets are available and can be obtained, if necessary.

4.4 Quality of Reference Data

The three studies that generated reference data were conducted using CFR Part 792 or CFR 160 GLPs.

4.5 Availability of Human Data

Relevant human data exist for each of the substances tested in the reference data studies. Human data were not used in generating the reference data.

4.6 Reference Data for the Computer Simulations

The computer simulations did not utilize any specific *in vivo* data; instead, the simulations encompassed the range of possible LD50 values and slopes as noted in the U.S. EPA's Office of Pesticides database.

4.7 Data Considerations

4.7.1 Data on Slopes and LD50 Values

A comparison of dose-response slope estimates for OECD TG 401 data using rats (29 substances from van den Heuvel et al., 1987) and U.S. EPA avian data (135 Office of Pesticides avian studies) is provided below in **Table 4-2**.

Table 4-2 Comparison of Dose-Response Slope Estimates for OECD TG 401 Rat Data (van den Heuvel et al., 1987) and U.S. EPA Avian Data (135 Office of Pesticides Avian Studies)

Slope	Number of substances (percent)	
	van den Heuvel	Avian
< 2.5	1 (3.4)	14 (10.4)
2.5 -6.0	11 (37.9)	77 (57.0)
> 6.0	17 (58.6)	44 (32.6)
	29	135

4.7.2 Avian Acute Toxicity and Slope Data

The avian data provided below are for registered pesticide active ingredients from the Environmental Fate and Effects Division (EFED) of the U.S. EPA database. The database file, called “bird_slopes”, contains only those studies for which a slope was recorded. Only 135 out of a total of 919 studies have reported slopes. Reasons for the slope not being reported include: (1) the study was a limit test, conducted at only a single dose level; (2) the study did not yield at least two doses with mortality between 0% and 100%, which is the minimal requirement of the analytical program (TOXANAL) U.S. EPA uses to calculate a

probit slope; (3) the study was conducted at dose levels either too high or too low; (4) mortality failed to follow a dose-response pattern; or (5) the slope was not calculated or recorded (common with older studies). It should be noted that studies with steeper slopes would likely not have a slope calculated for reason (2). Therefore, there may be a bias in the data in that steep slope values may be missing more frequently than shallow slope values.

Description of Field Names

CHEMICAL	Chemical common name
SHAUGHNESSEY	U.S. EPA identification number for active ingredient (Shaughnessey number)
USEPATTERN	Class of pesticide based on target organism (Ex. “insecticide”)
COMMONNAME	Species common name
TGL	Indicates if the toxicity value is “>” or “<”
TOXICITY	LD50 value in mg/kg
TOXLEVEL	Unit of toxicity value (MGK=mg/kg)
CL	95% confidence limit for LD50 estimate
CURVESLOPE	Probit slope estimate
EPAIDENT	U.S. EPA identification number for the study (MRID)

4.7.3 Data from Six Completed OECD TG 401 Studies

Summarized outcomes from six studies on five pesticides carried out according to OECD TG 401 are provided in this BRD. Issues relating to the analysis of pesticide data were the impetus for reexamining the performance of all alternative guidelines under various circumstances (i.e., shallow slopes). The data are tabulated giving proportion responding at each dose level, along with any estimates of LD50, slope, and associated confidence intervals, as well as the calculation method(s) cited by the study investigators. These data were cited in an U.S. EPA Office of Pesticide Programs study with confidential substance identity.

Compound 1: shallow dose response

Dose (mg/kg)	Males	Females
25 (prelim.)	0/2	0/2
100 (prelim.)	2/2	0/2
50	0/5	0/5
80	2/5	2/5
126	4/5	4/5
200	5/5	4/5
"LD50(95%CI)"	92(64-128)	103(73-141)

Using Finney's method for probits (1978), the male and female estimated slope is 5.5 (i.e., 1.4 with log transformation of dose), compared to a combined data estimated slope of 5.4 [i.e., 1.4 with log transformation of dose; LD50(95%CI) = 97(76-122)] (Finney, 1971).

Compound 2: shallow dose response

Dose (mg/kg)	Males	Females
987	0/5	0/5
1481	0/5	0/5
2222	3/5	3/5
3333	4/5	5/5
5000	5/5	not run
0	0/5	0/5
"LD50(95%CI)"	2314(1790-2990)	2132(1748-2600)

Using Weil (1952), the estimated LD50 and confidence intervals for combined male and female data was 2221 (1869-2639) mg/kg.

Compound 3: shallow dose response

Dose (mg/kg)	Males	Females
4000	0/5	0/5
4500	0/5	4/5
4800	0/5	5/5
5050	3/5	5/5
5200	2/5	not run
"LD50(95%CI)"	5150(4940- 5380)	4380(4210- 4560)

Using Litchfield and Wilcoxon (1949), the LD50 and confidence intervals for combined male and female data was 4810(4550-5080) mg/kg.

Compound 4: shallow dose response

Dose (mg/kg)	Males	Females
1	0/5	0/5
2	1/5	1/5
3	4/5	5/5
5	4/5	5/5
10	5/5	5/5
"LD50(95%CI)"	2.7(1.8-4.0)	2.7(1.8-4.2)

Using Litchfield and Wilcoxon (1949), the slope [(0.5)log(LD84/LD16)] was 0.23 for males and 0.15 for females, using the definition for compound 5.

Compound 5: variable dose response

Dose (mg/kg)	Males	Females
130	0/6	0/6
250	0/6	0/6
500	1/6	0/6
1000	0/6	3/6
2000	5/6	6/6
4000	6/6	6/6
"LD50(95%CI)"	1414(927-2598)	1000(733-1364)

Using Thompson and Weil (Biometrics 8:51-54) per C. Stephan (1978) the slope [(0.5)log(LD84/LD16)] was 4.1 for males and 3.8 for females.

Compound 6: steep dose response

Dose (mg/kg)	Males	Females
294/192	0/5	0/5
429/235	3/5	4/5
552/294	4/5	4/5
"LD50(95%CI)"	435(302-581)	234(183-296)

The calculation method is unspecified. However, a computer program of C.E. Stephan (1982) resulted in a slope of 10.6 for males and 13.4 for females.

5.0 Test Method Data and Results

There have been three studies in which data obtained using the UDP are compared with data obtained using OECD TG 401. A list of the substances tested in each study is provided in **Table 5-1**. In the Bruce (1987) and Bonnyns et al. (1988) studies, the OECD TG 401 data were generated simultaneously with the UDP data. In the Yam et al. (1991) study, the OECD TG 401 data were taken from a validation study for FDP (van den Heuvel et al., 1990) and little is known about the differences between animals and substances in the two studies.

5.1 *In Vivo* Data Using the UDP

5.1.1 Bruce (1987) Study

In the Bruce (1987) study, 10 substances were tested using a dose-spacing factor of 1.4 for OECD TG 401 tests and 1.3 for the UDP tests. For OECD TG 401, the animals were dosed simultaneously and observed for 14 days. For the UDP, the animals were dosed sequentially at least 24 hours apart and observed for 7 days. The stopping rule was that four animals were tested after the first reversal of outcome. The LD50 values for these substances ranged from 0.39 to 22 mg/kg and all calculated LD50 values for the two methods were within a factor of 1.4, well with the range observed in inter- and intra-laboratory variation studies (See **Section 7.0**).

5.1.2 Bonnyns et al. (1988) Study

In the Bonnyns et al. (1988) study, the UDP dose-spacing factor was 1.3 and five animals were tested after the first reversal. The selected substances affected different organs as follows:

barium acetate	heart
allyl heptanoate	central nervous system
barbital	central nervous system
coumarine	homeostasis
diquat	kidney

The published LD50 values ranged between 200 and 2000 mg/kg. All calculated LD50 values for the two methods were within a factor of 1.9, well within the range observed in inter- and intra-laboratory studies (See **Section 7.0**). Both OECD TG 401 and the UDP tests would have classified all substances as harmful.

5.1.3 Yam et al. (1991) Study

In the Yam et al. (1991) study, ten substances were tested in the UDP using a dose-spacing factor of 1.3 and the stopping rule was to test four animals after the first reversal. Animals were dosed sequentially, separated by 24 hours. The substances were also tested using the FDP by using five males and five females starting at one of the fixed dose levels. The animals weighed between 190 and 300 g, were fasted for 16 to 20 hours prior to dosing, and were observed for 14 days. The UDP LD50 data were compared to OECD TG 401 LD50 data of van den Heuvel et al. (1990). The OECD TG 401 data were generated in a single laboratory using the 1981 OECD guideline rather than the 1987 guideline, but no details as to strain, age, or weight of the animals were provided. The absolute ratio of each set of LD50 values for the UDP and OECD TG 401 were within a factor of 1.9, except for mercuric chloride where the ratio was 13. It is not clear why this discrepancy was present for mercuric chloride; it may be related to the purity/batch of the substance, solubility, weight or age of the animals, or other possible sources of variation as the OECD TG 401 data were taken from van den Heuvel et al. (1990). Additionally, one of the data points could represent an outlier. It should be noted that data in RTECS indicate that the LD50 for mercuric chloride is considerably less than 160 mg/kg.

Table 5-1 Substances and Results for the UDP Validation Studies

Test Substance	UDP LD50 (mg/kg)
Bruce (1987)	
Ingredient	22,400
Laundry detergent	11,090
Ingredient	>10,100
Shampoo	8,700
Dishwashing detergent	5,700
Flavor	4,120
Laundry detergent	4,020
Laundry detergent	3,520
Caffeine	421
Potassium hydroxide	388
Bonnyns et al. (1988)	
Diquat	1,022
Allyl heptanoate	582
Barbital	581
Coumarine	517
Barium acetate	302
Yam et al. (1991)	
p-Dichlorobenzene	2,495
Tetrachlorvinphos	2,208
4-Aminophenol	1,557
Acetanilide	1,107
Na pentachlorophenate	425
Piperidene	337
Fentin hydroxide	152
Nicotine	70
Na arsenite	53
Mercuric chloride	12

In the three validation studies involving the UDP, the resulting estimate of the LD50 was compared to an LD50 generated using OECD TG 401. The Revised UDP utilizes the same methodology as the UDP except in the dose-spacing factor and the stopping rules. On this basis, these studies can be applied to the validation of the Revised UDP. There was excellent concordance between OECD TG 401 and the UDP data for all 25 substances, except for mercuric chloride. The LD50 values ranged from 0.05 to 22 mg/kg and several chemical classes were represented.

6.0 Test Method Performance

The performance characteristics of the UDP and the Revised UDP can be evaluated using four criteria:

1. the point estimate of the LD50 as compared with OECD TG 401 data;
2. the estimation of the slope of the dose-response curve for mortality and the confidence interval for the LD50 as compared to OECD TG 401 data;
3. the hazard classification as compared to the hazard classification using OECD TG 401 data; and
4. the number of animals used in the study as compared to OECD TG 401.

6.1 *In Vivo* Validation Studies

In **Table 6-1**, the results of three *in vivo* validation studies involving OECD TG 401 and the UDP are provided along with the ratio of the LD50 values for the two methods. For all 25 substances, the average ratio of the LD50 values for the two methods is 1.76. If mercuric chloride is not included, the average ratio is 1.28. The LD50 using the Revised UDP was the higher value for 15 of the 25 substances and was the lower value for the remaining 10 substances. These data indicate that the two methods provide essentially the same point estimate of the LD50 for the substances tested. The single exception is mercuric chloride. Without access to the data for the OECD TG 401 LD50 values in the van den Heuvel (1990) study, it is impossible to determine whether significant differences (e.g., age or weight of the animals or purity of the test substance) between the two studies may have affected the outcome. In the Bruce (1987) and the Bonnyns et al. (1988) studies, the same laboratory determined the LD50 values using both OECD TG 401 and the UDP.

A comparison of rat oral LD50 data with estimated human lethality data is given in **Table 6-2**. The average ratio of the UDP LD50 to the lower estimate of human lethality is a factor of 46. This factor compares well with the safety factor of 100 often applied in risk assessment procedures to derive a safe level for humans while utilizing animal data. These data also illustrate and support the conservative approach of using safety factors in human risk assessment. On this basis, the UDP provides suitable data for risk assessment purposes and probabilistic modeling.

Table 6-1 Validation Studies for the UDP

Test Substance	LD50 (mg/kg)		Absolute Ratio of LD50 values
	OECD TG 401	**UDP**	
Bruce (1987)			
Ingredient	>10,000	>10,100	1.01
Laundry detergent	4,040	3,520	1.15
Ingredient	>20,000	22,400	1.12
Laundry detergent	3,510	4,020	1.15
Laundry detergent	10,110	11,090	1.10
Shampoo	9,280	8,700	1.07
Flavor	3,490	4,120	1.18
Caffeine	344	421	1.22
Potassium hydroxide	273	388	1.42
Dishwashing detergent	5,560	5,700	1.03
Bonnyns et al. (1988)			
Barium acetate	571	302	1.89
Barbital	404	581	1.44
Coumarine	470	517	1.10
Allyl heptanoate	991	582	1.70
Diquat	1,036	1,022	1.01
Yam et al. (1991)			
Nicotine	71	70	1.01
Na pentachlorophenate	309	425	1.38
Na arsenite	48	53	1.10
p-Dichlorobenzene	>2,000	2,495	1.25
Fentin hydroxide	119	152	1.28
Acetanilide	1,893	1,107	1.71
Tetrachlorvinphos	>2,000	2,208	1.10
Piperidene	488	337	1.45
Mercuric chloride	160	12	13.3
4-Aminophenol	>3,000	1,557	1.93
Average Ratio			1.76
Average Ratio (without mercuric chloride)			1.28

Table 6-2 UDP Study Substances with Human Oral Lethality Data

	UDP Rat LD50 (mg/kg)	OECD TG 401 Rat LD50 (mg/kg)	Dosage for 60 kg person* (mg/kg)
Bruce (1987)			
Caffeine	421	344	50 - 167
Bonnyns et al. (1988)			
Barbital	581	404	100 - 167
Diquat	1,022	1,036	67 - 100
Yam et al. (1991)			
Nicotine	70	71†	0.67 - 1.0
Sodium Arsenite	53	48†	1 - 20
Fentin Hydroxide	152	119†	1.17
Acetanilide	1,107	1,893†	0.83 - 8.33
Mercuric Chloride	12	160†	8.33
4-Aminophenol	1,557	>3000†	16.7

* Data from the Hazardous Substances Data Bank, National Library of Medicine (May 2000)
† Data from van den Heuvel et al. (1990)

6.2 Computer Simulation Validation of the Revised UDP

The Revised UDP is a statistical sampling technique designed to determine the mean and variance of the population of a test species. The Revised UDP has not been validated in *in vivo* studies; however, the current UDP has been validated against OECD TG 401 using *in vivo* studies. Because the Revised UDP involves only a change in statistical sampling technique, its performance cannot easily be determined using *in vivo* studies. Since computer simulations are more appropriate, the Revised UDP has been validated using this approach (see U.S. EPA Documents 5 and 6 - original BRD **Appendix C**, final report **Appendices K and L**, respectively).

6.2.1 Rationale for Statistical Approach for the Revised UDP

Acute oral toxicity tests provide quantal data because the result in any animal can be only one of two possibilities – either the animal lives or it dies. In evaluating a statistical method, the question will be, "How well does the method predict the mean and variance of the population based on a small sample taken from that population?" Consider an experiment to determine how often a flipped coin will come up heads or tails. Clearly the results of a single trial would be insufficient to determine the correct answer; even several trials would fail to provide the correct answer. Instead, the trials must be repeated over and over to determine how often the sampling technique will predict the correct answer.

6.2.2 How the Computer Simulations Work

The simulations are meant to represent all possible types of response configurations anticipated under the assumed conditions. To simulate an experiment, the following details should be known: the starting dose level; the underlying distribution of tolerances which is characterized by the LD50 and the slope of the dose-response curve; hazard classification; boundary doses; rules for handling boundary doses; and stopping rules. Additional information is needed for slope estimation experiments. By simulating

experiments under a set of assumed conditions, the distribution of possible outcomes can be characterized. The simulations take into account the variety of possible outcomes and the probabilities with which they are observed. In some cases, simulations are not necessary because distributional results can be used to determine test procedure performance.

For the Revised UDP, one experiment is simulated at a time and the LD50 is estimated. A total of 1000 to 5000 simulation experiments are conducted for each experimental design. This number of simulations is sufficient to achieve good representation of all of the experimental results likely to occur. The distribution of the LD50 estimates is then summarized and the 5^{th} and 95^{th} percentiles are reported.

The simulations are aimed at evaluating all of the permutations possible for the multiple experiments and do not provide the permutations possible for any one animal. If a given dose has 30% expected mortality, then on the average, in simulated experiments, that dose would produce lethality 30% of the time. However, as with any sample from a larger population, for any given set of animals receiving that dose, it should not be expected that exactly three of these ten animals (30%) would die.

6.2.3 Validation Using Computer Simulations

During a recent OECD evaluation of acute oral tests, all currently accepted designs were shown by simulation techniques to have poor ability to estimate the LD50 of the underlying population under two conditions: 1) when the dose-response curve is shallow and 2) when the starting dose level for the test is far from the actual LD50 (see U.S. EPA Document 1A – original BRD **Appendix C**, currently **Section 1.1.4** of this revised BRD). To determine if improvements in the sampling technique can be made to improve the ability of the Revised UDP to correctly estimate the LD50, simulations have been conducted (see U.S. EPA Documents 5 and 6 – original BRD **Appendix C**, final report **Appendix K and L**, respectively). Using simulations, the Revised UDP has a greater chance than the current UDP of placing the estimated LD50 near the mean of the underlying population, even when the starting dose level is inappropriate (**Table 6-1)**. This type of comparison would be impossible using actual animal tests, since no determination could be made regarding which small sample tested is providing the correct estimate of the underlying population and which sample is incorrect.

Instead, using LD50 data generated in past studies, a series of assumptions as to the slope, true LD50, and the starting dose level have been used to evaluate the Revised UDP as a statistical sampling technique. Using these assumed values, the UDP has been simulated to evaluate how well it estimates the true LD50 and slope using the various assumed values. The assumed values have been treated as though they are the mean and variance of the population. When both the mean and variance of the population are known, it is possible, using a computer, to simulate the generation of a random sequence of responses. Using this method, the computer can simulate the results from repeatedly taking small samples from a much larger population. The population is sampled in such a way that the results from the small sample have the best chance of correctly estimating the mean and variance of the entire population. By using a series of such simulations, it is possible to test how often the Revised UDP will accurately estimate the mean and variance or standard deviation of the population.

Animal testing is not only unnecessary, but is without value in determining the validity of the new statistical design. The characteristics of the test animal and the test methodology remain unchanged from the current UDP. Assay variability has previously been characterized and deemed acceptable by both the United States and international regulatory community. Thus, computer simulations provide the most suitable approach for evaluating changes in dose spacing and the decision criteria on estimating the LD50.

6.3 Results of Computer Simulations

Simulations and calculations have been conducted to explore the performance of the Revised UDP (see U.S. EPA Document 5 – original BRD **Appendix C**, final report **Appendix K**). Computer simulations have been used to optimize the protocol. The simulations have examined the spacing of doses, the efficiency of animal usage, starting dose level, assumed slope, and certain other factors. Simulations have also been used to examine the effects of steep and shallow slopes and the effects of the starting dose level being far from the LD50.

The UDP, as adopted, is designed to efficiently determine the LD50; to accomplish this task, a value for the slope and an estimate of the LD50, based on information available for the test substance, must be assumed. Nevertheless, the UDP does an excellent job of determining the LD50 except for substances with a shallow slope or in cases where the starting dose level is far from the "true" LD50. The U.S. EPA and other regulatory agencies need the slope of the dose-response curve and the confidence interval of the LD50 for certain substances for probabilistic modeling and risk assessment purposes.

The primary test in the Revised UDP is identical to the current UDP except for the dose-spacing factor, stopping rule, and other improvements. This procedure has been shown to efficiently estimate the LD50. The areas of improvement as evaluated via computer simulations are described below. Most of the changes evident in the Revised UDP involve the Supplemental Test and have been implemented to improve the estimation of the slope of the dose-response curve and the calculation of confidence interval of the LD50.

6.3.1 Dose-Spacing Factor

A discussion of the dose-spacing factor requires knowledge of slope and variance. The standard deviation for a data set is designated as *sigma* (σ) and *sigma* is the inverse of the slope of the dose-response curve; thus, a *sigma* of 0.5 corresponds to a slope of 2. *Sigma* is a measure the spread of the data around the center point in a lognormal bell-shaped curve (i.e., around the LD50). The method is optimized when the slope of the dose-response curve for the substance is near the assumed slope (the default spacing factor of 3.2 is optimized for a slope of 2). With the large spacing factor, the performance of the method is unaffected by the starting dose level, although the number of animals used will increase if the starting dose level is far from the LD50. For a shallow slope, the method is more likely to provide a correct estimate if the starting dose level is closer to the LD50. For a steep slope, the method provides a good estimate even if the starting dose level is far from the LD50 because the first reversal will be close to the LD50. For a shallow slope, the first reversal may occur far from the LD50 resulting in a bias toward the starting dose level. Thus, the probability of an early reversal (far from the LD50) depends on the slope, not the starting dose level.

The dose spacing in the current UDP is 1.3d, where d is the previous dose. This spacing corresponds to a slope value of 8 in the dose-response curve and a *sigma* of 0.125 in the normal curve of animal responses to the substance in a test for lethality. Simulations of the values for the LD50 calculated using the current UDP demonstrate that performance is optimum when the starting dose level is very close to the true LD50 and the assumed or assigned *sigma* is small and/or close to the true *sigma*. In fact, simulations show that the method works well for "true" *sigma* values < 0.25 (i.e., the median value estimated for LD50 is very close to the true LD50) and the 90% ratio (difference between 5^{th} and 95^{th} percentile predictions) of LD50 is relatively small (i.e., < 3). The probability of an early first reversal in test outcome depends on the distance of the initial dose from the true LD50.

If the starting dose level diverges significantly from the true LD50 and the spacing factor is 1.3d, the number of animals utilized to reach the LD50 can be excessive. When the starting dose level is far from

the true LD50 and the slope is shallow, a bias is introduced in the median value of the estimated LD50; in these cases, the bias is toward the starting dose level. When *sigma* is larger than the spacing factor, the spread of estimated LD50 increases. Simulations show that under these conditions, the 95/5% ratio may be highly variable and range from one or two orders of magnitude. For a spacing factor of 1.3d, shallow slopes do not increase animal usage, instead, the test terminates early because the first reversal is far from the LD50. However, steep slopes may cause an increase in animal usage if the starting dose level is far from the LD50 because it may take several doses to reach the lethal range for the substance when the spacing factor is small.

To reduce this inefficiency, consideration was given to changing the dose-spacing factor. After a number of simulation trials, it was found that use of a larger dose step size, namely 3.2d (or 0.5 log d), improved the efficiency of animal usage. In addition, when simulation experiments were performed with a 3.2d step size and calculations of LD50 used an assumed *sigma* value of 0.5 (corresponding to a slope of 2), the bias was minimized or eliminated in the median value of estimated LD50. However, there was only a slight improvement in the precision or the spread of estimated LD50 values (i.e., the 95/5% ratio). For substances with very shallow slopes or a large spread (*sigma* = 1.25), a bias in median value of LD50 reappears and the 95/5% ratio increases, but the problems are not as severe as with the smaller (1.3d) dose spacing.

A comparison of the median estimated LD50 (based on 1000 runs) and the number of animals used for dose-spacing factor of 1.3 and 3.2 is provided in U.S. EPA Document 5 (original BRD **Appendix C**, final report **Appendix K**). By increasing the spacing of doses, the efficiency of animal usage is improved and certain other characteristics are optimized in many simulations. The LD50 estimate using a spacing factor of 1.3 is very close to the actual LD50 for simulations using a steep slope; however, animal usage can be as high as 21. While the LD50 using a spacing factor of 3.2 is below the actual LD50, it never requires more than 10 animals. For moderate and shallow slopes, the spacing factor of 3.2 results in LD50 estimates that are more accurate and uses fewer animals than for LD50 estimates using the 1.3 spacing factor.

6.3.2 Use of a Stopping Rule

In cases where the slope of the dose-response curve is shallow, it may take many animals to determine an accurate LD50. If the test stops with four animals after the first reversal of outcome as is the case for the current UDP, the estimate of the LD50 is not very accurate; therefore, a stopping rule is needed to eliminate this inaccuracy. To obtain an accurate LD50, the test must be extended to include more animals when evaluating substances with a shallow slope. The stopping rule allows an accurate estimate of the LD50 while limiting the total number of animals to 15. If the slope is steep, the stopping rule has been designed to allow the test to stop at four animals after the first reversal. Based on the low percentage of substances with a shallow slope, the stopping rule will not increase animal usage for a majority of test substances. Five stopping rules have been considered as follows:

1. Based on fixed nominal size -- testing four additional animals after the first reversal; if a reversal is observed at the second dose level, the nominal size will be six.
2. Based on the number of reversals -- testing stops after five reversals; under the most favorable conditions (each dose level after the first resulting in a reversal), the number of necessary animals would be six.
3. Based on the convergence of estimators of the LD50 -- two estimators of the LD50 are the maximum likelihood estimate and the geometric average dose; testing stops when the ratio of the two estimators falls below 2 or other preassigned factor.
4. Based on a likelihood ratio with optimized slope -- values close to the geometric mean carry more weight than values far from the geometric mean; weight is determined using the likelihood ratio.

5. Based on a likelihood ratio with default slope -- identical to stopping rule #4 except a default slope is used, reducing the complexity of the calculations.

As stated above, stopping rule #1 does not work for shallow slopes. U.S. EPA Document 6 (original BRD **Appendix C**, final report **Appendix L**) provides a comparison of the number of animals used for each of the stopping rules with slopes varying from 0.5 to 8.3. Data are presented for starting dose levels of 0.1 LD50, LD50, and 100 LD50. On the basis of these data, stopping rules #1, #3, and #4 were not considered further.

The final stopping rule criteria are as follows:

1. The upper bound is reached and three consecutive animals survive at that bound or the lower bound is reached and three consecutive animals die at that bound.
2. The next animal to be tested would be the 7th and each surviving animal has been followed by a death and vice versa (i.e., five reversals occur in six animals dosed).
3. Beginning with the fourth animal after the first reversal (which may be as early as the 7th animal), three measures (likelihood estimates) of the test progress are compared using two ratios. If the first measure is at least two-and-one-half times both of the other measures (i.e., both ratios are at least 2.5), testing stops (see Appendix III in U.S. EPA Document 1B – original BRD **Appendix C**, final report **Appendix G**)

6.3.3 Other Considerations

6.3.3.1 Bounding of the Range of Test Dose Levels

The UDP has been revised so that test dose levels are bounded below by 1 mg/kg and above by 2000 or 5000 mg/kg. The features of the current algorithm (see U.S. EPA Document 5 - original BRD **Appendix C**, final report **Appendix K**) are the identification of a finite set of testable doses and a modification of the dose-spacing factor.

6.3.3.2 Stopping at the Bound Dose, "Out-of-Bound" Estimates (The Limit Test)

Testing stops if there is a sequence of three survivals at the designated upper limit dose level or a sequence of three deaths at the designated lower limit dose level. In those cases, the finding from the study is that the LD50 is outside the testable range (e.g., below 1 mg/kg or above 2000 or 5000 mg/kg). When the LD50 is calculated to be greater than 2000 or 5000 mg/kg, the experimenter would not use the point estimate of the LD50, but would merely conclude that the LD50 is above the upper limit dose level.

6.3.3.3 Performance Indices and Other Statistics Reported

The performance indices have been extended by including the percent of estimates "within a factor of 2" of the true LD50. The index is denoted PF2, standing for *P*ercentage with *F*actor-of- *2* accuracy. The index combines bias and precision.

When calculating measures of bias or spread, "out-of-bound" estimates are replaced with the nearest bound value (1 or 5000).

6.3.3.4 Maximum Number of Animals

The maximum number of animals tested has been set at 15. When 25 was used as the maximum number of animals, the number of animals tested was inflated in some situations even when the initial test dose was reasonable. Results using 15 animals were not markedly different from those using 25 animals.

6.3.3.5 Simulated Outlier Scenario

Due to concern regarding whether the simulation models adequately characterize the range of events occurring in actual lab situations, an "outlier scenario" has been simulated as follows: the initial test was assumed to be below the true LD50 (here 750 mg/kg) by a factor of 10 or 100 and the first animal tested was assumed to respond, regardless of the probability of response calculated from the probit model. The idea is that such an event could result from background mortality, mishandling, or administration of an incorrect dose level. When dealing with data which include an outlier, there is practically no chance for the nominal number (n = 6) stopping rule to provide a reasonable estimate of the LD50. This inability suggests that the stopping rule based on a nominal number of animals should be abandoned. The use of flexible-*n* stopping rules (e.g., based on the number of reversions or based on the maximum likelihood using a default slope) provided an appreciably higher probability of reasonable results as shown in U.S. EPA Document 5 (original BRD **Appendix C**, final report **Appendix K**).

6.4 Calculation of the Slope and Confidence Interval

A number of computer simulations have tracked the calculation of the slope depending on the assumed slope, the starting dose level, and the true LD50. These data are shown in U.S. EPA Document 6 (original BRD **Appendix C**, final report **Appendix L**). Two methods have been considered for calculation of the slope and confidence interval. One utilizes the UDP in the Supplemental Test and involves a multiple sequence dosing procedure in which three of four runs are conducted simultaneously. The second method (Group Method) is a modification of OECD TG 401 for the Supplemental Test.

6.4.1 Multiple Sequence Dosing

A number of variations of multiple sequence dosing have been simulated. In all cases, the LD50 is determined first. Then, three or four UDP tests are run in parallel beginning at slightly different starting dose levels. Each of these runs is complete when the first animal dies. The individual data for all runs, including the initial LD50 run, are then combined and used in a probit analysis to estimate the LD50 and slope of the dose-response curve. Data from computer simulations for this procedure are provided in U.S. EPA Document 6 (original BRD **Appendix C**, final report **Appendix L**). The number of animals used is greater than in the Primary Test, but only one animal per run (three or four total) should be killed by the test substance in the Supplemental Test.

6.4.2 Group Method Dosing

This method involves dosing groups of ten or more animals at established lethality points (e.g., LD10, LD16, LD84) derived from the dose-response curve. Data for this procedure are given in U.S. EPA Document 6, Part B (original BRD **Appendix C**, final report **Appendix L**). The group method labeled "Best Estimate" provides better results, but utilizes 30 animals not including those required for the LD50 determination (an additional seven animals for the LD50 determination). The group method works fairly well for steep slopes, but generally uses more animals than OECD TG 401 (37 animals plus seven animals for the LD50 determination).

6.5 Hazard Classification

All three of the *in vivo* validation studies resulted in the estimation of the LD50 for the substances studied; a direct comparison of the UDP to the OECD TG 401 in toxic classification is shown in **Table 6-3**. For the Bruce (1987) and the Bonnyns et al. (1988) studies, there is 100% agreement between the current UDP and OECD TG 401 in the classification of the tested substances. The Yam et al. (1991)

study, the FDP was conducted along with the UDP and the results were compared with the published results of van den Heuvel et al. (1990). The UDP gave the same classification as OECD TG 401 for eight of the ten substances tested. For the remaining substances, the UDP provided a more conservative classification. The FDP resulted in the same classification as OECD TG 401 for seven of the ten substances tested, was less risk averse for two substances, and was more risk averse for the other substance. When compared to the FDP, the UDP gave the same classification for eight of the ten substances and was more conservative for the other two substances (mercuric chloride and 4-aminophenol). A comparison of the results for FDP, ATC, and UDP is provided in **Table 6-4**. Overall, the UDP gave the same classification as OECD TG 401 for 92% of the substances tested and was more conservative (higher classification) for the remaining 8% of the substances tested.

Table 6-3 Toxic Classification

Test Substance	Toxic Classification		
	OECD TG 401	**UDP**	**FDP**
Bruce (1987)			
Ingredient	Unclassified	Unclassified	ND
Laundry detergent	Unclassified	Unclassified	ND
Ingredient	Unclassified	Unclassified	ND
Laundry detergent	Unclassified	Unclassified	ND
Laundry detergent	Unclassified	Unclassified	ND
Shampoo	Unclassified	Unclassified	ND
Flavor	Unclassified	Unclassified	ND
Caffeine	Harmful	Harmful	ND
Potassium hydroxide	Harmful	Harmful	ND
Dishwashing detergent	Unclassified	Unclassified	ND
Bonnyns et al. (1988)			
Barium acetate	Harmful	Harmful	ND
Barbital	Harmful	Harmful	ND
Coumarine	Harmful	Harmful	ND
Allyl heptanoate	Harmful	Harmful	ND
Diquat	Harmful	Harmful	ND
Yam et al. (1991)			
Nicotine	Toxic	Toxic	Toxic
Na pentachlorophenate	Harmful	Harmful	Harmful
Na arsenite	Toxic	Toxic	Toxic
p-Dichlorobenzene	Unclassified	Unclassified	Unclassified
Fentin hydroxide	Toxic	Toxic	Harmful
Acetanilide	Harmful	Harmful	Unclassified
Tetrachlorvinphos	Unclassified	Unclassified	Unclassified
Piperidene	Harmful	Harmful	Harmful
Mercuric chloride	Toxic	Very Toxic	Toxic
4-Aminophenol	Unclassified	Harmful	Harmful

VT = Very Toxic = LD50 $\leq$ 50 mg/kg; T = Toxic = LD50 > 50 mg/kg but $\leq$ 500 mg/kg;
H = Harmful = LD50 > 500 mg/kg but $\leq$ 2000 mg/kg; U = Unclassified = LD50 > 2000 mg/kg
ND = no data

Table 6-4 Comparison of the FDP, the ATC, and the UDP

OECD Test Alternative	Number of Test Substances	Number of Test Comparisons	Alternative Test Hazard Classification Compared to That of Standard Test (%)			Reference
			Same Hazard	Greater Hazard	Lesser Hazard	
FDP	41	41	75.6	4.9	19.5	van den Heuvel et al., 1987
	20	414	80.2	3.5	16.3	van den Heuvel et al., 1990
ATC	30	179	86	9.0	5.0	Schlede et al., 1992
	20	175	86	5.3	8.7	Schlede et al., 1995
UDP	25	25	92.0	8.0	0	Lipnick et al., 1995

7.0 Test Method Reliability (Repeatability/Reproducibility)

There are no known *in vivo* data on the reliability and repeatability of the Revised UDP. The current UDP has been shown to perform well when compared to OECD TG 401 (see **Section 6.0**). The OECD agreed when approving the UDP that the dosing method and observations were identical to OECD TG 401 and the ATCM, therefore, the inter- and intra-laboratory variability should also be identical. Data are presented for the repeatability and reproducibility acute oral toxicity studies. Using computer simulations, the repeatability and reproducibility of the Revised UDP has led to an optimized protocol.

7.1 Inter-laboratory Reproducibility for Acute Oral Toxicity Studies

In 1964, Griffith studied inter-laboratory variation in determining the acute oral LD50. Four substances were tested at six contract or industrial toxicity testing laboratories. Four laboratories utilized male and female Sprague-Dawley rats weighing between 200 and 300 g and two laboratories used male rats only. Four laboratories fasted the rats before dosing, whereas two laboratories did not fast the rats. The laboratories were free to decide how to prepare the doses and when a vehicle should be used. Five laboratories used water and one used corn oil. All substances were delivered to the laboratory as coded substances and all doses were administered via oral gavage. A total of four different statistical methods were used to calculate the LD50.

The ratio of the highest LD50 value to the lowest LD50 value ranged from 2.0 for sodium bicarbonate to 2.8 for sodium alkyl benzene sulfonate. The results for each substance are given in **Table 7-1**. For laboratories using the same concentration of the test substance in water, the resulting LD50 values were less variable. Dosing in corn oil seemed to lessen the toxic effects of the three substances administered in a vehicle, at least when the concentration in corn oil was the same as the concentration in water. Despite all of the differences in the acute oral toxicity protocol for these four substances, the LD50 values were all within a factor of 2.8.

In 1967, Weil and Wright completed an inter-laboratory comparison of eight laboratories studying the acute oral toxicity of 10 substances. Each laboratory conducted the test using three protocols. The first or standardized protocol specified the dose-spacing factor, the strain, weight, and number of rats, the rat diet,

and required overnight fasting of the animals. The second protocol was identical to the first except the laboratory could choose the strain of rat. The third protocol was not directed in any way (i.e., the laboratory conducted the test according to their standard procedures).

Using a standardized protocol, the ratio of the highest LD50 to the lowest LD50 for nine substances ranged from 1.5 to 2.8 as shown in **Table 7-2**. For the 10th substance, the ratio was 5.0. Some of the variability resulted from one laboratory inadvertently utilizing specific pathogen free rats instead of conventional stock rats as specified in the protocol. For that laboratory, the LD50 values were relatively higher when compared to the other laboratories.

Table 7-1 Ratio of Highest to Lowest Inter-Laboratory LD50 values from Griffith (1964)

Test Substance	Highest LD50	Lowest LD50	Ratio
Sodium Bicarbonate	8.29	4.22	1.96
Akylbenzene sulfonate	5.82	2.05	2.84
Granular detergent	7.92	3.56	2.60
Liquid detergent	16.15	7.25	2.23

Table 7-2 Inter-Laboratory LD50 values from Weil and Wright (1967)

	Substance									
Laboratory	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	2.24	2.59	0.71	5.66	0.21	3.25	8.00	6.73	0.77	6.50
2	2.12	1.50	0.42	5.60	0.20	2.38	8.48	4.06	1.23	4.24
3	2.46	2.80	0.28	5.90	0.21	4.92	9.90	8.91	1.97	8.12
4	1.62	1.87	0.71	4.92	0.27	4.92	7.46	7.46	1.23	2.83
5	2.46	1.23	0.54	4.29	0.13	2.83	6.50	2.83	0.81	3.36
6	2.26	1.97	0.57	4.53	0.17	3.94	6.86	9.05	0.70	4.85
7	1.54	1.54	0.34	3.54	0.13	4.06	8.12	14.1	1.17	5.45
8	2.14	1.19	0.71	4.24	0.16	4.00	9.85	5.04	1.29	3.57
Absolute LD50 Ratio	1.6	2.4	2.5	1.7	2.0	2.1	1.5	5.0	2.8	2.8

The results using the second protocol were almost identical to the results for the standardized protocol; the results using the third protocol were much more variable. For these third protocol studies, nonfasted rats and more mature rats (weighing between 220 and 310 g) resulted in significant differences in the LD50 values.

7.2 Intra-Laboratory Repeatability for Acute Lethality Studies

In 1966, Weil and coworkers reported results for an intra-laboratory study of the acute oral toxicity of 26 substances. The LD50 values were determined for almost all substances in 11 of 12 consecutive years. Each test utilized nonfasted rats (predominantly males) weighing between 90 and 120 g. Over the 12 years, six strains of rats were used and eleven technicians were involved with dosing. The substances were administered neat, in water, in corn oil, or in Tergitol®.

The ratio of the highest LD50 to the lowest LD50 value for each substance ranged from 1.33 for dipropylene glycol to 3.18 for monoethanolamine. The results for all 26 substances are provided in **Table 7-3**. Considering the variations in strains of rat, varying use of a vehicle, and different technicians, the acute oral toxicity test is quite reproducible.

In 1967, Weil and Wright reported the results of an acute oral toxicity study conducted in eight laboratories using ten different substances. Each laboratory conducted the test using three protocols. By comparing the results for the three protocols for each laboratory, an indication of intra-laboratory variation was ascertained. The specific LD50 data were not provided, but the data were reported using a ranking procedure. Using a relative rank procedure based on the sum of ranks for all 10 substances, essentially no differences were noted in the three protocols as the sum of ranks were 15, 15, and 17, respectively, as shown in **Table 7-4**.

Table 7-3 Intra-Laboratory Repeatability from Weil et al. (1966)

Test Substance	LD50 Ratio (High/Low)
Mesityl oxide	2.00
2,4-Pentane dione	1.63
2-Ethyl butyric acid	3.02
Isophorone	2.96
Diethanolamine	2.19
Morpholine	1.74
Monoethanolamine	3.18
Butyl cellosolve	2.11
2-Ethyl hexanoic acid	2.19
2-Ethyl hexanol	2.11
Methyl cellosolve	1.65
n-Butanol	2.43
Diethyl carbitol	2.28
2-Ethylhexenediol	3.15
Diisobutyl ketone	2.25
Diacetone alcohol	1.50
Butyl carbitol	2.72
Triethanolamine	2.05
Ethylene glycol	2.00
Methyl carbitol	1.56
Carbitol	1.96
UCON LB-400	2.79
Dipropylene glycol	1.33
Diethylene glycol	1.74
Triethylene glycol	1.92
Propylene glycol	1.52

Table 7-4 Relative Rank of Sum of Ranks for LD50 values (Weil and Wright, 1967)

	Laboratory								
Procedure	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Sum**
I	3	1	2	2.5	1	3	1.5	1	15
II	2	2	1	2.5	2	1	1.5	3	15
III	1	3	3	1	3	2	3	2	17

7.3 Other Studies

Zbinden and Flury-Roversi (1981) reviewed acute oral toxicity data from the open literature and noted many factors that may affect the determination of the LD50 including:

animal species
age of the animals
weight of the animals
sex of the animals
genetic influence (strain differences)
animal health
diet
food deprivation
dosing procedure
ambient temperature
housing conditions
seasonal variations
humidity
light/dark cycle
noise
weather (barometric pressure)
technician training
acclimation period

All of these factors are important and over time the protocol has become standardized in an attempt to minimize variability. After Zbinden and Flury-Roversi (1981) noted these factors affecting variability, they claimed the LD50 test was unreliable because the open literature shows values ranging from 3.66 to 11.89 fold. It should be noted that the data producing high variability were not generated using a standardized protocol (e.g., the weight of the male rats varied from 52 to 400 g); had the data been generated using a standard protocol, they likely would not have varied beyond a factor of three, as observed in the studies summarized above.

Based on inspection of LD50 data available from RTECS or other reference texts and databases, the LD50 reported for several species and multiple strains using differing protocols varies by a factor of 10 or more. Such a compilation is not adequate to evaluate inter- or intra-laboratory variation.

7.4 The Need for Additional Repeatability/Reproducibility Studies

Reference acute oral toxicity data were obtained from inter- and intra-laboratory studies using protocols predating OECD TG 401. It is clear from these results that the protocols for acute oral toxicity studies needed to be standardized if the results for various studies are to be compared. OECD TG 401 is standardized and the results in inter- and intra-laboratory studies show that the method provides an estimate of the true LD50 within a factor of approximately three. As OECD TG 401 has been considered the classical method for many years, new or alternative methods should yield results comparable to those obtained using this protocol.

7.5 Inter-Laboratory Reproducibility Studies Using the FDP and the ATC

Two multi-laboratory international studies have generated data regarding the inter-laboratory reproducibility of two acute toxicity methods. In the first study, van den Heuvel et al. (1990) reported the results of 33 laboratories in 11 countries studying 20 coded substances using the FDP. With participation from 33 laboratories, one laboratory advised on preparation and distribution of the 20 substances, a second laboratory performed a classical LD50 test on each substance, and the remaining 31 laboratories conducted the FDP. The laboratories performing the FDP were free to choose the strain of rat; 21 used Sprague-Dawley rats, 9 used Wistar rats, and one used Fischer 344 rats. The age of rats at study initiation was from 8 to 12 weeks and their weight was ±20% of the mean. The exact strain, age, and weight used in each study were not provided. Animals were dosed at 5, 50, 500, or 2000 mg/kg and the results were matched with the then current European Commission (EC) classification scheme. The reproducibility of the FDP is illustrated in **Table 7-5.**

Of 516 comparisons, the authors reported 414 (80.2%) of the FDP classifications were the same as the LD50 test. For 84 comparisons (16.3%), the FDP underclassified the substances and for 18 comparisons (3.5%), the FDP overclassified the substances. Fentin hydroxide, 2-chloroethanol, and 4-aminophenol were underclassified by 69%, 27%, and 35% of the testing laboratories, respectively. 1-Phenyl-2-thiourea was overclassified by 46% of the testing laboratories. The authors stated that the variability of the results for 1-phenyl-2-thiourea was probably due to solubility problems. For fentin hydroxide, wide variations were due in part to strain and weight differences in the rats; the Fischer 344 rats used by one laboratory were reported to be twice as large as the other strains. This variation equates to large differences in age because Fischer 344 rats are usually smaller than Sprague-Dawley or Wistar rats of the same age. The results for 4-aminophenol and 2-chloroethanol were not readily explained. According to the authors, the FDP produces "consistent results that are not substantially affected by inter-laboratory variation."

In the second study, Schlede et al. (1995) reported the results of nine laboratories in five countries studying 20 coded substances using the ATC. Six laboratories used Sprague-Dawley rats, and three laboratories used Wistar rats. No specifications as to age or weight were given except that the weights for all rats used were reported to be ±20% of the mean at study initiation for each laboratory. Based on a comparison with LD50 data (selected from various sources in the open literature), eight of the 20 substances were classified correctly by all laboratories reporting data. The reliability of the ATC is illustrated in **Table 7-6.**

Of 173 comparisons, 136 (79%) of the ATC classifications were the same for all laboratories reporting data. Indomethacin, *N*-phenylthiourea, and bis(tributyltin)oxide were underclassified by 56%, 56%, and 78% of the testing laboratories, respectively. Cadmium chloride was overclassified by 67% of the testing laboratories. No explanation was provided for these deviations. According to the authors, the ATC is "a reliable alternative to the LD50 test."

Despite the variability due to strain, age, and weight of rats, the FDP and the ATC were reasonably consistent for all of the substances tested (only three substances spanned three classes). These two international studies support the overall reproducibility of *in vivo* acute toxicity data and would suggest that there is no need for additional *in vivo* inter-laboratory validation studies for the UDP (see U.S. EPA Document 13; original BRD **Appendix C**, final report **Appendix J-1**).

Table 7-5 Inter-Laboratory Reproducibility of FDP (van den Heuvel et al., 1990)

Substance	LD50 (mg/kg)	Number of Labs Classifying (n=26)*		
		Correctly	Over	Under
Class 3 (0 - 25 mg/kg)†				
Aldicarb (10%)	3.2-5.0	22		
Class 2 (25 – 200 mg/kg)				
Phenyl mercury acetate	37	24	2	
Sodium arsenite	48	25		1
2-Chloroethanol	60	19		7
Nicotine	71	23		3
Fentin hydroxide	119	8		18
1-Phenyl-2-thiourea	126-400	12	12	2
Mercuric chloride	160	25		1
Class 1 (200 – 2000 mg/kg)				
Sodium pentachlorophenate	309	25	1	
Piperidine	488	24	2	
Resourcinol	489	25		1
Ferrocene	1260-2000	3		23
Acetanilide	1893	4		22
Class 0 (2000 – ∞ mg/kg)				
p-Dichlorobenzene	>2000	26		
Quercetin dihydrate	>2000	26		
Tetrachloevinphos	>2000	25	1	
Naphthalene	>2000	26		
Acetonitrile	>2000	22	4	
Dimethyl formamide	>2000	26		
4-Aminophenol	>3000	17	9	
Totals (n=516)		407	31	78

*Correctly = predicted same hazard classification as OECD TG 401; Over = predicted greater hazard than OECD TG 401; Under = predicted lesser hazard than OECD TG 401

†Actual doses utilized were 5, 50, 500, or 2000 mg/kg

Table 7-6 Inter-Laboratory Reproducibility of ATC (Schlede et al., 1995)

Substance	LD50 (mg/kg)	Number of Labs Classifying (n=9)*		
		Correctly	Over	Under
Class 3 (0 – 25 mg/kg)				
Aldicarb	1	9		
Parathion	4	9		
N-Phenylthiourea	9	4		5
Thiosemicarbazide	12	9		
Indomethacin	13	4		5
Class 2 (25 – 200 mg/kg)				
Mercuric oxide	29	8	1	
Sodium arsenite	38	8	1	
Bis(tributyltin)oxide	147	2		7
Acrylamide	163	8		1
Class 1 (200 – 2000 mg/kg)				
Cadmium chloride	237	3	6	
Caffeine	270	8	1	
Aniline	822	9		
Ferrocene	1280	9		
Sodium salicylate	1601	6		
Acetanilide	1689	5		3
Class 0 (2000 - ∞ mg/kg)				
Acetonitrile	2515	5	3	
Butylated hydroxyanisole	2853	5	3	
N,*N*-Dimethylformamide	4604	7	1	
Quercetin dihydrate	>2000	9		
Ethylene glycol	6336	9		
Totals (n=173)		136	16	21

*Correctly = predicted same hazard classification as OECD TG 401; Over = predicted greater hazard than OECD TG 401; Under = predicted lesser hazard than OECD TG 401

8.0 Test Method Data Quality

8.1 Adherence to Good Laboratory Practices (GLPs)

The studies of Bruce (1987) and Yam et al. (1991) were conducted under CFR Part 792 GLPs. The Bonnyns et al. (1988) study was conducted in Belgium under GLPs of the European Community.

8.2 Results of Data Quality Audits

The QA audit report for the Bruce (1987) study was not available; however, the signed report regarding the conduct of the study according to GLPs was provided. For the Yam et al. (1991) study, the laboratory report including all observations, body weights, and pathology were provided. Individual data sheets for one of the substances were also provided. The QA audit report was not available, but from the data provided, no serious deviations from GLPs were noted. QA audits, study reports, and animal data were not available for the Bonnyns et al. (1988) study or the van den Heuvel et al. (1990) study (the source of the OECD TG 401 data for the Bonnyns study).

8.3 Impact of GLP Deviations and/or Data Audit Non-Compliance

A review of the Bruce (1987) and the Yam et al. (1991) studies did not reveal any discrepancies that would have significantly altered the general conclusions of the study reports.

9.0 Other Scientific Reports and Reviews

9.1 Availability of Additional UDP Data

The only other known toxicity data using the UDP are the unpublished data from the Netherlands (see original BRD **Appendix D;** this appendix was not included in this final report). These data are quite different in that birds were used and were dosed two at a time, resulting in the use of many birds (some sixty animals per study).

9.2 Inhalation Testing and the UDP

Inhalation toxicity testing is more complex than oral or dermal toxicity testing. The purpose of an acute inhalation toxicity study is to provide an assessment and evaluation of the toxic characteristics of an inhalable substance, such as gases, volatile substances, or aerosols/particulates. It also provides information of possible health hazards to a human if exposed via the inhalation route. An acute inhalation toxicity study determines the median lethal concentration (LC50) and its statistical limits and slope using a single exposure duration (usually of 4 hours) and a 14-day post-exposure observation period. Data from an acute study can serve as a basis for classification and labeling; it is also an initial step in establishing a dosage regimen in subchronic and other studies, and might provide additional information on the mode of toxic action of a substance (Technical Committee of the Inhalation Specialty Section, 1992).

Current U.S. EPA guidance indicates that at least five animals of the same sex should be used at each test concentration (Gross and Vocci, 1988; Gross, 1989). After completion of the study in one sex, at least one group of animals of the other sex is exposed to characterize any differential sensitivity to the test substance. The U.S. EPA encourages the use of fewer animals if justified in individual circumstances. Where adequate information is available to demonstrate that animals of the sex tested are markedly more sensitive, testing of the other sex is not required. Where appropriate, a Limit Test may be considered. In the Limit Test, a single group of five males and five females is exposed to 2 mg/L for four hours. In

situations where this concentration is not possible due to the physical properties of the test substance, the animals are exposed to the maximum attainable concentration. If no lethality is observed, no further testing for acute inhalation toxicity is needed. If compound-related mortality results, further study may need to be considered.

Testing one animal at a time, in either a nose only or a whole body exposure chamber, would greatly increase the cost of the assay. The increase in study cost results primarily from the additional chamber time needed, as well as the additional analyses for concentration and particle size required for each run. Study costs would also be increased because the exposure chamber will be unavailable for a different study until the UDP is completed, since only then could the generation system be cleaned and prepared for another test substance. Additionally, from a practical standpoint, compared to simultaneously exposing all animals to the same test concentration, exposing single animals at different times to exactly the same test concentration is more difficult. Thus, it does not appear currently that using a sequential dosing procedures such as the UDP for inhalation toxicity testing is a viable alternative.

9.3 Other Acute Toxicity Methodology

One method worth considering as an alternative to the UPD is the method of Weil (1983). In this method, four groups of three or four animals are dosed using a dose-spacing factor of 2 and the LD50 and slope are calculated using the moving-average method. Using a dose-spacing factor of 1.26 or 2.0, Weil et al. (1953) showed that groups of three or four animals yield an estimate of the LD50 equivalent to that determined using groups of ten animals; thus, with 12 to 16 animals, the LD50, slope, and confidence interval could be determined in a single study. The moving-average method can accommodate dose groups that have 0% or 100% kills. Calculating the slope using probit analysis requires the use of many more animals. In a comparison of 35 pairs of slopes determined using probit analysis and the moving-average method, the correlation coefficient was 0.85. If the dosing is performed in sequence, three dose levels may be sufficient for the study, thereby requiring only 9 to 12 animals total.

Weil (1975) summarized the results of 490 probit analyses for acute oral tests; these summaries generated a median slope of 7.8. Only 8 of 490 had a slope of 2 or less and more than 50 had a slope of 16 or greater, ranging up to a slope of 60; this fact confirms that relatively few test substances have a slope of 2 or less. It also indicates that even for a relatively simple one-dose test, the slope of the dose-response curve for different test substances is quite variable. The uncertainty of the slope in each assay is large compared to the relatively low degree of uncertainty of the LD50. Even with this uncertainty, the slope estimate is critical for risk assessment purposes and probabilistic modeling.

10.0 Animal Welfare Considerations

10.1 Refinement to Address Animal Pain and Suffering

In the Yam et al. (1991) study, the number of toxic signs and deaths in the UDP and OECD TG 401 were compared. The results clearly show that in the UDP, the incidence and severity of pain and suffering were reduced when compared to OECD TG 401. The Revised UDP specifically refers to the OECD Guidance 19 (original BRD **Appendix B;** this appendix is not appended to this final report) on humane endpoints and handling of moribund animals. The use of this guidance document in the training of technicians is key to the refinement process.

10.2 Reduction in Animal Usage

The 1981, OECD TG 401 utilized 50 or more animals to calculate the LD50, slope, and confidence interval. The 1987 revision of OECD TG 401 reduced that number to 20 to 30 animals. The Revised UDP is designed to use 6 to 15 animals in the LD50 determination. The utilization of animals is compared in **Table 10-1** for the three validation studies. A summary table comparing the Revised UDP to OECD TG 401 is presented in **Table 10-2**.

Table 10-1 Animal Usage in OECD TG 401 and the UDP

	Number of animals	
	OECD TG 401	**UDP**
Bruce (1987)	370	68
Bonnyns et al. (1988)	150	40
Yam et al. (1991)	260	75
TOTALS	780	183

The UDP utilized only 23% of the animals used in OECD TG 401, yet the estimated LD50 values were in good agreement. For the LD50 determination, the Revised UDP will use the same or fewer numbers of animals (usually females) as is used by the current UDP.

Table 10-2 Summary Table of Acute Oral Toxicity Tests
(Assume nothing is known about test substance)

	TG 401 (1981)	**TG 401** (1987)	**TG 425** (1998)	**Revised UDP** (2001)
Range-finding study (RFS)	yes	yes	NA	NA
# doses	$\geq$ 3[a]	$\geq$ 3 [a]		
# animals/dose	5	3		
males/females	both	one		
total animals	30+	9+		
duration[b]	7 days	7 days		
LD50 Estimate	yes	yes	yes	yes
# animals/dose	5/sex	5	1	1
# dose levels	3-6	3+1[c]	2-13[d]	2-6[e]
males/females	both	1/confirm	females	females
total animals	30-60	20 [c]	6-18 [d]	6-10 [e]
starting dose	from RFS	from RFS	100 mg/kg	175 mg/kg
duration[b]	14 days	21 days	22-39 days	26-35 days
Totals for RFS plus LD50 Estimate				
# animals	**60 - 90+**	**29+**	**6 -18**	**6 - 10**
duration[b]	**21-28 days**	**28-35 days**	**22-39 days**	**26-35 days**
Slope Estimate	possible[f]	possible[f]	NA	yes (Supplemental Test)
# runs				4
# doses/run	0-2	0-2		1-4
# animals/dose	5/sex	5		1
total animals	0-10	0-10		4 –16[g]
duration[b]	0-14 days	0-14 days		14-18 days
Combined Totals (LD50 and Slope estimates)				
# animals	**60 -100+**	**29 - 39+**	**NA**	**16 - 25**
duration[b]	**21-42 days**	**28-49 days**	**NA**	**42-53 days**

[a] minimum of three doses; more if lethality range not bracketed in the first three doses.

[b] assume dosing on Monday – Friday only; duration for all tests includes a 14-day post-dosing observation period.

[c] three doses tested in first sex plus one dose tested in second sex.

[d] starting at 100 mg/kg with a spacing factor of 1.3, 13 dose treatments could occur prior to the first reversal (e.g., the first death at 2000 mg/kg in this example) – 100, 130, 169, 220, 286, 371, 483, 627, 816, 1060, 1380, 1790, and 2000 mg/kg. The total number of animals used would then be 13 plus the 4 after the first reversal or 17 animals. If the animal dosed at 2000 mg/kg lived, then a Limit Test would be conducted (up to 5 more animals for a total of 18 animals).

[e] starting at 175 mg/kg with a spacing factor of 3.2, six dose treatments could occur prior to the first reversal (e.g., the first animal to survive in this example was at a dose of 1.0 mg/kg) – 175, 55, 17.5, 5.5, 1.75, and 1.0 mg/kg. The total number of animal would then be 6 plus the 4 after the first reversal or 10 animals. If the animals dosed at 1.0 mg/kg died, then a lower Limit Test would be conducted (up to 4 more animals, also a total of 10 animals).

[f] slope estimation requires three dose groups for each sex with partial kills; if not achieved in the LD50 determination, then one or more dose groups may be required.

[g] if the first animal in each run dies, then the total is four animals; if death is not observed until the 4th animal in each run, then the total is 16 animals.

10.3 Replacement of the Acute Oral Toxicity Test

Concern has been expressed about the reliability and usefulness of acute oral toxicity tests (Zbinden and Flury-Roversi, 1981). Recently, for humane reasons, increasing interest and support have been given to the use of *in vitro* cytotoxicity methods. Recent advances in *in vitro* cytotoxicity methodology, especially through the Multicentre Evaluation of *In Vitro* Cytotoxicity (MEIC) Program and through validation studies conducted at the Center for Documentation and Evaluation of Alternative Methods to Animal Experiments (ZEBET), have been reported (Ekwall, 1999; Halle, 1998). However, *in vitro* cytotoxicity tests have not yet been validated as a replacement for acute oral toxicity tests. It is possible that such tests could be used to determine the starting dose level in animal studies. An *In Vitro* Cytotoxicity Workshop, sponsored by ICCVAM, has been scheduled for October 17 - 19, 2000 in Crystal City, VA, U.S. to explore these issues.

11.0 Other Considerations

11.1 Gender Sensitivity

Several documents regarding sex sensitivity issues have been reviewed (see U.S. EPA Document 14 - original BRD **Appendix C**, final report **Appendix P**). Because data suggest that the female is more sensitive in the majority of instances, the use of females in the Revised UDP will result in a more protective number in risk assessment action and probabilistic modeling.

11.2 Equipment and Training

The equipment requirements for the Revised UDP are no different than for other acute oral toxicity studies, with the possible exception of the requirement of a computer. Cages, balances, analytical equipment as necessary to confirm the identity of the test substance, possibly waterbaths or mixers to dissolve the substance, dosing syringes, gavage catheters, and necropsy equipment are needed. The only special piece of equipment needed for this revised method is a computer to run a spreadsheet program and a means to run maximum likelihood estimates using an appropriate statistical program. It is anticipated that the stopping rule program will be made available in Excel® or another standard format to interested individuals via the OECD or U.S. EPA websites. A program could also be written, as described in the UDP guideline, by the investigator.

Training requirements are similar to any acute oral toxicity test with emphasis on recognizing animals in a moribund condition and other humane endpoints (see original BRD **Appendix B;** this appendix is not appended to this final report). Technicians must be trained to properly calculate, mix, and administer test substances to rats via oral gavage and trained to make and record observations in an acute toxicity study, including the gross necropsy. They should also be able to make decisions on when to sacrifice a terminally ill animal.

Staff must also be able to use the computer programs. A full description of how to use the stopping rule, with examples, is in the guideline. The use of the maximum likelihood method for calculating the LD50 is a standard statistical program and would require someone with appropriate experience. Dosing and observations are similar to any other acute toxicity protocol. It is important for all acute toxicity studies that the technicians running the studies be trained in making and recording observations correctly.

11.3 Costs Comparisons for TG 401 and UDP Studies

Three commercial toxicology laboratories were contacted regarding costs of conducting OECD TG 401 and OECD TG 425. The comparisons are given below.

Test	Laboratory 1	Laboratory 2	Laboratory 3
Range-Finding Study	$800	$950	$2,900
Limit Test	$2,000	$1,650	$2,900
TG 401 (3 dose levels)	$5,000	$3,600	$6,900
UDP			$6,900
Primary Test	$2,000	$3,300	
Limit Test	$2,000	$1,650	
Supplemental	$800/run	$300/animal	

For Laboratory 1, the cost for an OECD TG 401 study is $5,000. For the UDP, the cost would be $2,000 for the Primary Study plus $3,200 (four runs) for the Supplemental Test for a total of $5,200. Thus, the costs are essentially equal.

For Laboratory 2, the cost for the OECD TG 401 study is $950 plus $3,600 for three levels for a total of $4,550. For the UDP, the Primary Test is $3,300 plus $2,400 (four runs with 2 animals each) for a total of $5,700. In this laboratory, the UDP cost is slightly greater than that for TG 401.

For Laboratory 3, the cost of the OECD TG 401 study and the UDP study (Primary and Supplemental) are equal.

Overall, the cost of the UDP study appears to be essentially the same as for the OECD TG 401 study. However, as many laboratories are not experienced with the UDP, these costs estimates may be expected to change.

11.4 Time Comparisons for Conducting TG 401 and UDP Studies

The UDP will require approximately two additional weeks when compared to OECD TG 401. This added time is attributed to the sequential dosing of all animals at 48-hour intervals in each UDP run and to the fact that the Primary Test is completed prior to the start of the Supplemental Test. In terms of technician time, there is little difference between the two tests as suggested in the above cost analysis.

12.0 References

ASTM (American Society for Testing and Materials). 1987. Standard Test Method for Estimating Acute Oral Toxicity in Rats. ASTM E1163-87. In: Annual Book of ASTM Standards, Philadelphia.

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.

Bonnyns, E., M.P. Delcour, and A. Vral. 1988. Up-and-Down method as an alternative to the EC-method for acute toxicity testing. IHE Project No. 2153/88/11. Institute of Hygiene and Epidemiology, Ministry of Public Health and the Environment, Brussels. 33 pp.

Bruce, R.D. 1987. A confirmatory study for the up-and-down method for acute toxicity testing. Fundam. Appl. Toxicol. 8: 97-100.

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

Dixon, W.J. 1991. Design and analysis of quantal dose-response experiments (with emphasis on staircase designs). Dixon Statistical Associates, Los Angeles CA, USA.

Ekwall, B. 1999. Overview of the Final MEIC Results: II. The *in vitro*/*in vivo* evaluation, including the selection of a practical battery of cell tests for prediction of acute lethal blood concentrations in humans. Toxicol. *In Vitro* 13(4-5):665-673.

Finney, D.J. 1978. Statistical Methods in Biological Assay. Griffin, London.

Finney, D.J. 1971. Probit analysis. 3rd ed. Cambridge Univ. Press, Cambridge, U.K.

Griffith, J.F. 1964. Interlaboratory variations in the determination of acute oral LD50. Toxicol. Appl. Pharmacol. 6: 726-730.

Gross, S.B. April 18, 1989. Memorandum subject: comments on standard evaluation procedure. Inhalation Toxicology Testing (SEP/Inhalation).

Gross, S.B., and F.J. Vocci. August 1988. Hazard evaluation division standard evaluation procedure inhalation toxicity testing. U.S. EPA-540/09-88-101.

Halle, W. 1998. Toxizitätsprüfungen in zellkulturen für eine vorhersage der akuten toxizität (LD50) zur einsparung von tierversuchen. Life Sciences/ Lebens-wissenschaften, Volume 1, 94 pp., Jülich: Forschungszentrum Jülich.

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. Food Chem. Toxicol. 33: 223-231.

Litchfield, J.T., Jr. and F. Wilcoxon. 1949. A simplified method of evaluating dose-effect experiments. J. Pharmacol. Exp. Ther. 96:99-113.

Luttik, R. and T. Aldenberg. 1997. Extrapolation Factors for Small Samples of Pesticide Toxicity: A Special Focus on LD50 Values for Birds and Mammals. Environ. Toxicol. Chem. 16: 1785-1788.

Organisation for Economic Co-operation and Development (OECD). 1999. OECD Guideline for Testing Chemicals Revised 420: Acute Oral Toxicity - Fixed Dose Procedure. OECD, Paris. [updated in 2000]

OECD. 1999. OECD Guideline for Testing Chemicals Revised 423: Acute Oral Toxicity - Acute Toxic Class Method. OECD, Paris. [updated in 2000]

OECD. 1998. OECD Guideline for Testing Chemicals 425: Acute Oral Toxicity: Up-and-Down Procedure. OECD, Paris.

OECD. 1998. Harmonized Integrated Hazard Classification System for Human Health and Environmental Effects of Chemical Substances as endorsed by the 28th Joint Meeting of the Chemicals Committee and Working Party on Chemicals in November 1998, Part2, pg 11. http://www.oecd.org/ehs/Class/HCL6.htm

OECD. 1987. OECD Guideline for Testing Chemicals 401: Acute Oral Toxicity. OECD, Paris.

OECD. 1981. OECD Guideline for Testing Chemicals 401: Acute Oral Toxicity. OECD, Paris.

OPPTS Harmonized Test Guidelines. August 1998. Series 870 Health Effects, Volume I of III, Guideline OPPTS 870.1300 Acute Inhalation Toxicity.

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

Schlede, E., W. Diener, U. Mischke, and D. Kayser. 1994. OECD expert meeting: acute toxic class method. January 26-28, 1994, Berlin, Germany.

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

Schlede, E., U. Mischke, R. Roll, and D. Kayser. 1992. A national validation study of the acute toxic class method - an alternative to the LD50 test. Arch. Toxicol. 66: 455-470.

Sheehan P.J., A. Baril, P. Mineau and D. Paustenbach. 1995. "Predicting the Effects of Pesticides on Aquatic Systems and the Waterfowl that Use Them. Chapter 30 in "Fundamentals of Aquatic Toxicology". Second edition, G.M. Rand (Ed.), Taylor and Francis, North Palm Beach, Florida. pp. 827-857.

Technical Committee of the Inhalation Specialty Section. 1992. Society of Toxicology recommendations for the conduct of acute inhalation limit tests. Fundam. Appl. Toxicol. 18: 321-327.

Trevan, J.W. 1927. The error of determination of toxicity. Proc. Royal Soc. 101B: 483-514.

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. Food Chem. Toxicol. 28: 469-482.

van den Heuvel, 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.

Weil, C.S. 1983. Economical LD50 and slope determinations. Drug Chem. Toxicol. 6 :595-603.

Weil, C.S. 1975. Toxicology experimental design and conduct as measured by interlaboratory collaborative studies. J. Off. Anal. Chem. 58: 683-688.

Weil, C.S., and G.J. Wright. 1967. Intra- and interlaboratory comparative evaluation of a single oral test. Toxicol. Appl. Pharm. 11: 378-388.

Weil, C.S., C.P. Carpenter, and H.F. Smyth. 1953. The median effective dose. Ind. Hyg. Q. 14: 200-206.

Weil, C.S., C.P. Carpenter, J.S. West, and H.F. Smyth. 1966. Reproducibility of single oral dose toxicity testing. Am. Ind. Hyg. Assoc. J. 27: 483-487.

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. Food Chem. Toxicol. 29:259-263.

Zbinden, G., and M. Flury-Roversi. 1981. Significance of the LD50 test for the toxicological evaluation of chemical substances. Arch Toxicol. 47: 77-99.