

## Methods

### CAN WE REDUCE THE NUMBER OF FISH IN THE OECD ACUTE TOXICITY TEST?

HANS RUFLI\*<sup>†</sup> and TIMOTHY A. SPRINGER<sup>‡</sup>

<sup>†</sup>ecotoxsolutions, Basel, Switzerland

<sup>‡</sup>Wildlife International, Easton, Maryland, USA

(Submitted 27 July 2010; Returned for Revision 19 November 2010; Accepted 15 December 2010)

**Abstract**—OECD (Organisation for Economic Co-operation and Development) Guideline 203, Fish Acute Toxicity Test, states that the test should be performed using at least five concentrations in a geometric series with a separation factor not exceeding 2.2, with at least seven fish per concentration. However, the efficiency of this design can be questioned, because it often results in only one concentration that causes partial mortality (mortality  $>0\%$  and  $<100\%$ ). We performed Monte Carlo computer simulations to assess whether more efficient designs could allow reductions in fish use. Simulations indicated that testing with six fish per concentration could yield 50% lethal concentration (LC50) estimates of quality similar to those obtained using seven fish. Experts attending a workshop organized to consider this finding and to identify the best methods for reducing fish use concluded that significant reductions could best be achieved by modifying the test paradigm. They suggested initiating testing using a 96-h fish embryo test instead of juvenile fish to cover the range from the upper threshold concentration (the lowest 50% effective concentration [EC50] in existing algae and daphnia studies) to the highest concentration with no mortality. This would be followed by a confirmatory limit test with juvenile fish at the highest concentration with no mortality or by a full test with juvenile fish, if a point estimate of the LC50 is required. Environ. Toxicol. Chem. 2011;30:1006–1011. © 2011 SETAC

**Keywords**—Fish acute test    Concentration response slope    Concentration separation factor    Fish rangefinder    Fish embryo test

#### INTRODUCTION

The Organisation for Economic Co-operation and Development (OECD) Test Guideline (TG) 203, Fish Acute Toxicity Test [1], article 17, states that at least five concentrations with a separation factor preferably not exceeding 2.2 should be used and at least seven fish per concentration. However, it is not uncommon in fish tests to find just one concentration with a partial mortality (mortality  $>0$  and  $<100\%$ ), and, because the goal of performing a test according to TG 203 is to estimate the 50% lethal concentration (LC50), little information is gained from the multiple test concentrations with no or complete mortality. Indeed, the results of many fish acute tests performed for regulatory submission with low-toxicity chemicals must be expressed as a one-sided interval, such as  $LC50 > 100$  mg/L. These observations suggest that more efficient test designs might be able to provide high-quality LC50 estimates using fewer fish than required by TG 203. It is the aim of the present study to find alternative testing approaches that provide high-quality LC50 estimates using reduced numbers of fish.

#### MATERIALS AND METHODS

Results from a series of 523 LC50 studies (96 h) performed according to OECD TG 203 from 1990 to 2000 in a laboratory of Ciba-Geigy, Novartis, and Syngenta Crop Protection were compiled into a database, called the *industry laboratory database* in this paper. In addition, results from 4,010 studies performed according to the comparable EPA guidelines (OPPTS 850.1075 and FIFRA 72-1) were extracted from the U.S. EPA Office of Pesticide Programs (OPP) Ecotoxicity database (<http://www.ipmcenters.org/Ecotox/index.cfm>) for comparison. The patterns of mortality in the two databases

were examined to identify opportunities for design improvement, and the distributions of LC50 values and slopes observed in these studies were characterized to provide background information useful in modeling the performance of different experimental designs. Potential design improvements were evaluated through the use of Monte Carlo simulations that modeled the interaction of experimental designs with the mortality patterns typically observed in dose–response tests as well as the statistical process of estimating the LC50 from the mortality data.

The overall goal of the simulations was to attempt to find a design for the acute toxicity study that reduces the number of fish used while producing LC50 estimates as good as those from the present study design. The simulations were based on the concept that, for a given chemical, there is a concentration–response curve characterized by a true LC50 and a true slope that we attempt to estimate by observation of experimental results. Because of the random nature of responses (mortality) in experiments, estimates of LC50 and slope derived from the results of repeated tests under the same conditions will have distributions that are shaped by the true LC50, true slope, number of test animals, number of concentrations, and spacing of the concentrations. The Monte Carlo simulations consisted of repeating 50 simulated experiments for various combinations of these factors. The true LC50 and true slope used were representative values selected to cover the range of values common in the previously described databases. Values for the LC50 (16, 1.2, and 0.12 mg/L) and slope (13, 8, 4, and 2) were chosen as the basis for generation of mortality patterns in the simulations (Table 1). It was assumed that the concentration separation factor could not be less than 1.6 because of limitations in maintaining and measuring test concentrations precisely and should not be much greater if the goal of having at least two concentrations with partial mortality was to be achieved. Thus, all simulations were performed with a concentration separation factor of 1.6.

\* To whom correspondence may be addressed (ruffi@ecotoxsolutions.com).

Published online 20 January 2011 in Wiley Online Library (wileyonlinelibrary.com).

Table 1. Selected centile values for 50% lethal concentration (LC50) and concentration–response slope from the subset of studies with point estimates in the industry laboratory and U.S. EPA Office of Pesticide Programs (OPP) databases (LC50 values in  $\mu\text{g/L}$ )

Parameter	Centile	Industry laboratory	OPP	Selected for simulations
LC50	75	15,859	12,589	16,000
	50	5,012	1,259	1,200
	25	1,585	126	120
Slope	75	18	4.1	13
	50	13	3.1	8
	25	3.8	2.1	4
	10	2.1		2

It was also assumed that, in real-world testing, an initial informed guess at the LC50 would be used as the value around which the test concentrations would be distributed in a range-finding test and that the LC50 estimate from the rangefinder would be used as the center of the concentrations used in the definitive test. Some of the simulations performed examined the benefit of combining the rangefinding results with the results from the definitive test. The simulations included study designs consisting of rangefinders with one to four concentrations and two to five fish per concentration, and definitive tests with three to five concentrations and four to 10 fish per concentration that resulted in over 1,000 permutations of design and response curve characteristics (true slope, true LC50).

The quality of the LC50 estimates from the various experimental designs was quantified by characterizing the distributions of the LC50 values that were obtained for each combination of experimental design and response–curve characteristics. Specifically, the bias (distance of the median of LC50 estimates from the true LC50) and ratio of 95th to 5th centile of the distribution of LC50 estimates ( $R_{95/5}$ , a measure of precision or spread) was determined for each distribution. A set of statistical methods for calculation of the LC50 values for each simulated study was incorporated into the simulations. The set of methods was selected to mimic the procedures typically used in testing laboratories to handle both simple and problematic results as follows. If two test concentrations resulted in partial mortality in the groups of exposed animals, then classical probit maximum likelihood techniques [2] were used to estimate the LC50. If mortality increased from 0 to 100% as concentration was increased from one concentration to the next, then probit methods or the method of Litchfield and Wilcoxon [3] could not be used, and the LC50 was estimated as the midpoint on the log scale between the concentration with complete survival and the concentration with complete mortality. This is the maximum likelihood estimate for this concentration–response pattern. When neither of these concentration–response patterns was observed, but the concentration causing 50% mortality could be determined by interpolating between the responses in two test concentrations, the LC50 was estimated using Stephan's nonlinear interpolation method (C.E. Stephan, unpublished data). Nonlinear interpolation is used because standard maximum likelihood methods cannot be used for this mortality pattern. No LC50 estimate can be calculated if there is complete survival or complete mortality in all exposure concentrations, or if none of the concentration–response patterns described above is observed. Inclusion of methods for estimating LC50 values when probit analysis cannot be used is extremely important, because these methods must be used in a high percentage of studies.

A workshop was held to consider the results of the simulations as well as broader approaches to reduction of animal use that rely on fundamental changes in the process of obtaining acute toxicity data on fish, rather than simply modifying the number of concentrations, number of animals per concentration, or spacing between concentrations. The workshop was held in Basel, Switzerland, on 27–28 October, 2009, and was attended by 18 representatives of academia, industry, and public authorities from Europe and the United States. Subgroups developed concepts for a test strategy to reduce fish numbers further in the concentration range process, which was presented to the entire group. Proposals drew on advances in various areas of ecotoxicology testing, including the new OECD 223 sequential test design [4], fish embryo testing [5], and use of the upper threshold (UTC) concept of multispecies testing [6,7].

## RESULTS

### *The industry laboratory database*

The industry laboratory database contained results from tests with agrochemicals and a variety of other types of chemicals. Among the 523 studies in the database, 329 provided point estimates of the LC50, whereas 194 studies provided one-sided interval estimates of the LC50 (LC50 > limit test concentration). The subset of 329 tests with point estimates for the LC50 included 228 tests of agrochemicals (69% of 329) and 101 tests of nonagrochemicals (31% of 329). The subset of 194 studies reported as limit tests included tests of 103 agrochemicals and 91 nonagrochemicals, corresponding to 31 and 47% of 194, respectively.

The distributions of the LC50 values and the slopes of the concentration–effect relationships for agrochemicals and non-agrochemicals were similar. Partial mortality (mortality >0 and <100%) was observed in two or more concentrations in approximately 30% of the 172 agrochemical studies and in 14% of the 78 studies with nonagrochemicals for which sufficiently detailed data were available to make the determination of number of partial mortalities. Approximately 75% of all studies had fewer than two concentrations showing partial mortality. In the absence of two partial mortalities, estimates of the LC50 and the slope of the concentration–response curve are approximations, and the results from most test concentrations (0 and 100% mortality) contribute little information. These findings strongly suggest the need for better selection of the spacing of the test concentrations and also suggest that, if the placement of test concentrations is improved, the number of animals required for testing might be reduced. However, as discussed below (see scenario 4), spacing between concentrations is constrained by practicality, and it became clear that improvements in design performance could not be readily attained by changing this aspect of the experimental designs.

$\text{Log}_{10}(\text{LC50})$  values from the subset of studies with point estimates were approximately normally distributed with a median of 3.7 (antilog of 5,012  $\mu\text{g/L}$  on the original scale is 3.7). Neither the distribution of the slope estimates nor the  $\text{log}(\text{slope})$  values was normally distributed. The median of the slope estimates was 13, and there were two modes in the distribution (near 3 and 15). Selected centiles of the distributions of LC50 and slope values are presented in Table 1.

### *The OPP database*

The OPP database contains data from 4,010 acute studies with fish. Among the 2,013 studies categorized as core, 326 LC50 values were reported as greater than a stated value

(a one-sided interval estimate of the LC50) and 1,684 as discrete LC50 values. The Log<sub>10</sub> discrete LC50 values were approximately normally distributed, with a median of 3.1 (antilog of 1,259 µg/L on the original scale is 3.1).

Among the 4,010 acute fish studies in the database, only 652 (16%) had reported slope values and 469 (22%) of the 2,013 studies categorized as core. Therefore, it is very difficult to make quantitative statements about the distribution of slopes in these studies. Given this caveat, the median slope value was 6.5, with a single strong mode in the distribution near this value. Selected centiles of the distributions of LC50 and slope values are presented in Table 1.

#### Comparison of the databases

The distributions of the LC50 values in the industry laboratory and OPP databases were roughly similar, having unimodal distributions and median values of 5.0 and 1.3 mg/L, respectively. However, the distributions of slopes in the databases differed markedly, being bimodal with a median of 13 in the industry laboratory database and unimodal with a median of 3.1 in the OPP database. The differences between the databases indicate that care should be taken in accepting the values in any database as being representative, regardless of size of the data set. One consequence of these findings is that the simulations described in this paper had to be performed for a very wide range of slopes to ensure that likely concentration–response scenarios were evaluated.

#### Simulation studies

Given that the simulations included study designs consisting of rangefinders with one to four concentrations and two to five fish per concentration and definitive tests with three to five concentrations and four to 10 fish per concentration and that more than 1,000 permutations of design and response curve characteristics (true slope, true LC50) were evaluated, results of all the simulations cannot be presented. Fortunately, it is possible to focus on a limited subset of the simulations that resulted in LC50 estimates that were of quality similar to those derived using the standard test design. Four such designs are presented below (Table 2).

**Scenario 1.** Scenario 1 was 47 + 7: total of 47 fish in rangefinder and definitive test plus 7 in control. This scenario is based on performing a definitive test according to the standard experimental design described in OECD TG 203 consisting of five concentrations with seven fish each following a rangefinder consisting of three concentrations with four fish each. Three concentrations separated by a factor of 10 with four fish each was common practice in the industry laboratory cited. For REACH, the concentration rangefinding has been replaced by the step-down procedure after applying the UTC approach. However, rangefinding is still used, particularly for pesticides.

The concentration separation factor is 1.6. Studies performed according to this scenario resulted in little bias, with the ratio of the median LC50 estimates to true LC50 value ranging between 0.97 and 1.05. Indeed, bias was minimal in all of the simulations performed, regardless of model or type of concentration response. The precision (spread) of LC50 estimates, as measured by  $R_{95/5}$  (ratio of 95th to 5th centile of LC50 estimates from the simulations), proved to be the most sensitive measure of study quality. Values of  $R_{95/5}$  for the LC50 distributions obtained in the simulations varied from 1.2 to 6.4 (Fig. 1), depending on the slope of the underlying concentration–response curve and the true LC50 value. The greatest spread of LC50 estimates (lowest precision) was observed in the low-slope simulation scenarios. Note that in Figure 1 there are three points for simulations for each of the true slopes used for simulation. Each of these points represents the results for a different true LC50. However, the size of  $R_{95/5}$  was not related to the value of the true LC50, so for simplicity the values of the true LC50 are not indicated. The large variation in the relation between results from the different designs seen for the low-slope concentration–response curves reflects two factors. First, the small number of simulations performed causes the calculation of  $R_{95/5}$  to be sensitive to outliers. Second, extreme outlying LC50 values occur periodically when the slope of the concentration response is low (that is, the tolerance distribution is broad and sensitive to outliers). Thus, the three sets of simulations in which the true slope was 2 differ by relatively large amounts. Variation in  $R_{95/5}$  seen for slope 2 is not systematically related to the true LC50.

Minimizing bias and  $R_{95/5}$  are the quality goals that any design with reduced numbers of fish must meet. An experimental design that results in LC50 distributions with larger bias or  $R_{95/5}$  values is less robust than alternatives that result in less bias or smaller  $R_{95/5}$  values. Additional design goals related to these primary goals include ensuring that estimates of LC50s and concentration–response slopes are obtained in a high percentage of studies. The frequency of failing to obtain estimates of the LC50 and slope are dependent on both the model and the true underlying slope of the concentration–response curve. Under the standard design (scenario 1), the percentage of simulated experiments that failed to provide point estimates for the LC50 ranged from 0%, when the slope was 2, up to almost 20%, when the slope was 13. These failure percentages differ from those observed in the industry laboratory and OPP databases because the randomly chosen LC50 values used to generate concentration–response curves in the simulations were almost always below the 100 mg/L limit on test concentrations that is imposed by the test guideline, whereas it is apparently common for the true LC50 values for the chemicals in the databases to be above 100 mg/L.

**Scenario 2 (42 + 7 fish).** When simulated tests were performed according to the design described in the standard

Table 2. Parameter settings for simulations<sup>a</sup>

Scenario	Scenario	Merged	$k_{rf}$	$n_{rf}$	$k_{dt}$	$n_{dt}$	Spacing factor <sub>rf</sub>	Spacing factor <sub>dt</sub>
Two stages; current standard scenario	1	N	3, 4	3, 4, 5	5	7	10	1.6
Two stages; standard scenario but with $n_{dt}=6$	2	N	3, 4	3, 4, 5	5	6	10	1.6
Two stages; standard scenario but with $k_{dt}=4$	3	N	3, 4	2, 3, 5	4	7	10	1.6
Two stages; standard scenario but with $n_{dt}=5$	4	Y	3, 4	2, 3, 4, 5	5	5	10	1.6

<sup>a</sup>Two stages = rangefinder (rf) followed by definitive test (dt); merged = the data of the rangefinder and of the definitive test are merged;  $k_{rf}$  = number of concentrations used in the rangefinder;  $n_{rf}$  = number of fish used in the rangefinder;  $k_{dt}$  = number of concentrations used in the definitive test;  $n_{dt}$  = number of fish used in the definitive test; spacing factor<sub>rf</sub> = spacing factor of concentrations in rangefinder; spacing factor<sub>dt</sub> = spacing factor of concentrations in definitive test.

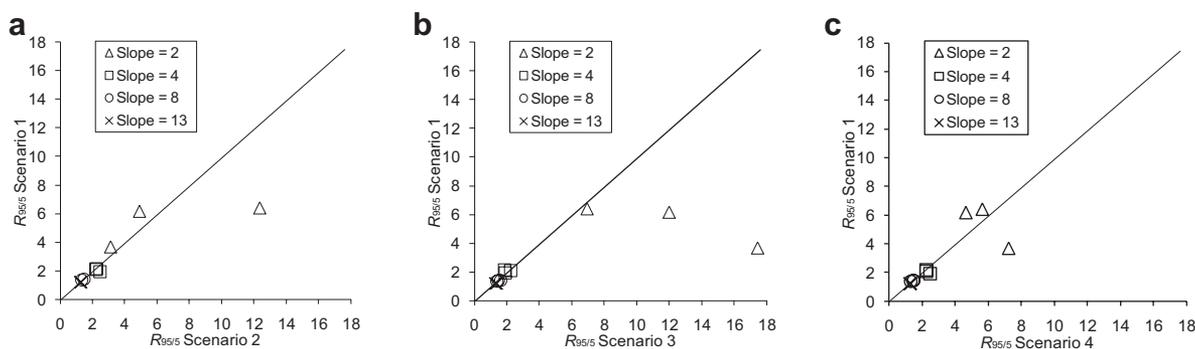


Fig. 1. Values of  $R_{95/5}$  (ratio of 95th to 5th centile of LC50 estimates) as a measure of precision of the LC50 distributions obtained in the simulations (lower values of  $R_{95/5}$  correspond to a higher precision, lower variability). (a) y Axis:  $R_{95/5}$  of scenario 1 (47 + 7 fish: 47 fish in rangefinder and definitive test plus 7 in control): rangefinder with three concentrations containing four fish each; definitive test consisting of five concentrations with seven fish each. x Axis:  $R_{95/5}$  of scenario 2 (42 + 7 fish): same as scenario 1, but definitive test with six instead of seven fish per concentration. (b) y Axis:  $R_{95/5}$  of scenario 1. x Axis:  $R_{95/5}$  of scenario 3 (40 + 7 fish): standard approach, but reducing the number of concentrations in the definitive test to four. (c) y Axis:  $R_{95/5}$  of scenario 1. x Axis:  $R_{95/5}$  of scenario 4 (37 + 7 fish): rangefinder with three concentrations containing four fish each; definitive test consisting of five concentrations with five fish each, pooling rangefinding and definitive test results. LC50 = median lethal concentration.

scenario (scenario 1), but with the definitive test being performed with six instead of seven fish per concentration (Table 2), the results were very similar to those obtained in scenario 1. The precision (spread) of LC50 estimates as measured by  $R_{95/5}$ , varied from 1.3 to 12.4 depending on the slope of the underlying concentration–response curve and the true LC50 value. The plots of the results from scenario 1 versus scenario 2 fall close to the line of perfect agreement except when the slope is less than 4. Therefore, it appears that, under the conditions simulated, using six fish per concentration would not lower the quality of the result compared with the present standard scenario, except in rare cases when the slope of the concentration–response curve is very low (e.g., less than 4), but would reduce the number of exposed fish by five per test. The percentage of simulated experiments that failed to provide point estimates for the LC50 was similar to that observed in scenario 1.

**Scenario 3 (40 + 7 fish).** If the standard approach was used with only four concentrations in the definitive test, the results did not deviate greatly from those of the standard scenario when slopes were 13, 8, and 4, but the precision of LC50 estimates was reduced when the slope was 2.

**Scenario 4 (37 + 7 fish).** The simulations under scenario 4 focus on the possibility that, if rangefinders were performed under conditions sufficiently similar to those of the definitive test (e.g. flow-through system for both the rangefinder and the definitive test), it might be possible to pool the results of the rangefinder and definitive tests to provide an improved estimate of the LC50. The results presented are for a rangefinder of three concentrations with four fish in each concentration, followed by a definitive test consisting of five concentrations with five fish each. Under these conditions,  $R_{95/5}$  was 1.3 to 7.2, which compares favorably with the standard design (scenario 1), in which  $R_{95/5}$  was 1.2 to 6.4. Simulations with further decreases in the number of test animals or test concentrations led to clear reductions of precision in LC50 estimates, even if data from the rangefinder and definitive test were pooled.

The initial impetus for the investigation reported here was the idea that the current OECD experimental design often leads to studies that have fewer than two concentrations with partial mortality. Because the desired method for estimating the LC50 for a study is the use of probit analysis by maximum likelihood methods, which requires two partial mortalities, it appears that many concentrations are essentially wasted. Our logic was that we might be able to reduce the number of wasted concentrations

and thereby reduce the number of test animals. The most obvious approach to reducing wastage in this way would be to reduce the separation factor between concentrations to ensure the occurrence of more partial mortalities. However, this is not possible because of practical limitations in aquatic testing, in which variations in test concentration resulting from dosing and analytical difficulties result in a nonmonotonic series of test concentrations, if concentration separations are less than about 1.6. Furthermore, the simulation results suggest that failure to obtain two or more partial mortalities does not lead to inaccurate LC50 estimates. Indeed, the simulations indicate that, over the long run, low-slope situations give rise to the greatest variability in LC50 estimates. However, when the slope is low, multiple partial mortalities are likely to be obtained. It is important to keep in mind that a low slope of concentration–response curve indicates high variability among responses, so the wide range of LC50 estimates obtained is understandable. Consequently, for an individual study, a high slope might make it impossible to assess the LC50 and uncertainty of the LC50 estimate using maximum likelihood methods. Paradoxically, over many repetitions of studies for a high slope chemical, the variability of the LC50 estimates (as indicated by  $R_{95/5}$ ) will be low, provided that the LC50 can be estimated by methods other than standard probit analysis. It is probably possible to achieve modest reductions in use of animals without reducing precision of the LC50 estimates in the high-slope situations. However, for any chemical that produces a low-slope concentration–response curve, even modest reductions in animal use are possible only if studies are designed so that rangefinding and definitive test results can be pooled to estimate the LC50. It is common practice to perform rangefinders under static conditions and the definitive test under flow-through conditions. Pooling could be performed only if current practices were modified so that rangefinding and definitive tests were performed under the same flow conditions. Unfortunately, the inconsistencies between the distributions of slopes in the historical databases make it difficult to sort out just how likely the low- and high-slope concentration–response curves actually are.

#### Alternative approaches to acute testing and test design

The results of the simulation studies clearly show that large reductions in animal use are likely to require a fundamentally different approach in the experimental design. Consideration of the simulation results in the workshop led to the conclusion that

efficiency in rangefinding is probably the key to reductions in animal use. Workshop participants eventually settled on strategies for reducing fish use in the acute test based on the use of the 96-h fish embryos for rangefinding [5] instead of juvenile fish. (The zebrafish fish embryo toxicity test is now generally considered to extend to 96 h and is classified as a nonanimal test according to the definitions of the European Union animal protection directive [8] because the embryo does not begin to feed freely until after this period [9]. This is also considered in the ongoing fish embryo toxicity test validation at OECD (S. Belanger, personal communication)). Furthermore, the rangefinding strategies thought to be most efficient involved starting tests with the embryos at the upper threshold concentration (UTC) corresponding to the lowest 50% effective concentration (EC50) value from algae and *Daphnia* tests with the chemical [6,7] if available or another reasonable starting concentration  $\leq 100$  mg/L based on the information available. An evaluation of 694 acute tests revealed that fish were the most sensitive in only 15.6% of these tests, whereas in 84.4%, the fish LC50 was  $\leq$  UTC [10]. If the embryo test shows no toxicity at this concentration, it is followed by an in vivo fish confirmatory test performed as a limit test, or as a full test if a concentration relationship is required. If the concentration is toxic, the embryo test is repeated, stepping down from the previous test concentration until there is no toxicity, or by a rangefinder using fish embryos, followed by the in vivo confirmatory limit test with fish (five fish in a single concentration and control for hazard classification [6], and seven for risk assessment [1]; end testing if there is no mortality).

## DISCUSSION

It was interesting to find different distributions in the two databases (industry laboratory database: bimodal, median of slopes 13; vs. OPP database: unimodal, median of slopes 6.5). The reasons are not known but might include different modes of action or artifacts such as from selection criteria for inclusion of tests in the databases. However, the present study did not analyze the individual data of the two databases regarding the influence of modes of action, nor selection criteria, to find the real cause.

A reduction to five fish per concentration in the definitive test (scenario 4) can be applied, if rangefinders were performed under conditions sufficiently similar to the definitive test (e.g., flow-through system for both the rangefinder and the definitive test). However, most rangefinders are conducted to gain basic information for testing, generally in a static system, even if the definitive test will be conducted under flow-through conditions. Thus, situations that apply for scenario 4 do not occur frequently in practice.

Stepping down in concentration until no mortality is observed is the method recommended in the threshold approach using juvenile fish. In the present study, this method is mentioned for the alternative approach using embryos. This corresponds to the outcome of the workshop. However, for a fast nonanimal test such as the 96-h embryo test, it might be more practical to conduct a full rangefinder compared with stepping down one concentration.

## CONCLUSIONS

The LC50 values and the slopes of the concentration–response relationships for agrochemicals and nonagrochemicals were similar. The proportion of limit tests was 31% of the

agrochemical and 47% of the nonagrochemical tests in the industry laboratory database.

Steep concentration–response slopes (median 13) were common in the database from the industry laboratory. This means that for high slopes above 8, the lowest possible concentration separation factor of 1.6 (in practice, separation factors cannot be much smaller than 1.6) is still too large to obtain two partial mortalities.

The OPP database contains a far lower percentage of studies with high slopes than does the industry laboratory database. The median of the slopes in the OPP database was 6.5. Therefore, for the studies in the OPP database, a concentration separation factor in the range of 1.6 to 2.1 would be adequate to get two partial mortalities in a much higher percentage of studies than would be likely in the industry laboratory database.

Unless the slope of the concentration–response curve is low, using six fish per concentration should yield LC50 estimates of quality similar to that obtained using the seven fish presently required by the OECD TG 203. Using five fish while maintaining the quality of LC50 estimates in tests of chemicals that produce low-slope concentration–response curves may be possible, if studies are designed in a similar way so that rangefinding and definitive test results can be pooled to estimate the LC50.

The resulting test strategies of the workshop were to use the 96-h fish embryo test proposed as an OECD Guideline for rangefinding instead of the juvenile fish test. Testing should be started at the UTC corresponding to the lowest EC value of algae and daphnia tests, if available. If there is no toxicity in the embryo test, it is followed by an in vivo fish confirmatory test performed as a limit test, or a full test if an LC50 value is required. Otherwise, the embryo test is repeated, stepping down the concentration until there is no toxicity, or by a rangefinder using the embryo test. When the no-toxicity concentration is found for the embryos, an in vivo confirmatory test with fish is performed.

The results of the project show ways of reducing the number of animals in the acute fish testing procedure. First, for steep concentration–response slopes, the simulations indicated that testing with six fish per concentration could yield LC50 estimates of quality similar to that obtained using seven fish, but, for low-slope conditions, the quality of LC50 values can be maintained while reducing fish numbers only if the mortality data from rangefinding and the definitive test can be combined to estimate the LC50. Second, reduction in animal use is possible by using juvenile fish only for a final confirmatory test, whereas any previous tests (rangefinders) are performed with fish embryos. To really save animals in practice, however, Guideline 203 has to be adapted (allowance of six fish per concentration instead of seven, and recommendation of the 96-h embryo test for initial rangefinding). For this purpose, a lead country has to submit a Standard Project Submission Form to the OECD.

*Acknowledgement*—The study was supported by the 3R Research Foundation, Switzerland, project 114/08. The authors are grateful for the valuable contributions of the participants in the workshop: M. Böttcher, T. Braunbeck, P. Chapman, J. Green, S. Jeram, R. Lange, G. Maak, U. Memmert, J. Navas, C. Polleichtner, C. Schafers, H. Segner, J. Straub, K. Thorpe, J. Wheeler, and S. Zok.

## REFERENCES

1. Organisation for Economic Co-operation and Development. 1992. Fish, Acute Toxicity Test. OECD Guideline 203. Paris, France.

2. Finney DJ. 1964. *Statistical Method in Biological Assay*. Charles Griffin, London, United Kingdom, pp 1–668.
3. Litchfield JT, Wilcoxon F. 1949. A simplified method of evaluating dose–effect experiments. *J Pharmacol Exp Ther* 96:99–113.
4. Organisation for Economic Co-operation and Development. 2007. Avian Acute Oral Toxicity Test. Proposal for a new OECD Guideline 223. Paris, France.
5. Organisation for Economic Co-operation and Development. 2006. 48h-Fish Embryo Toxicity Test (FET). Draft Proposal for a New OECD Test Guideline. Paris, France.
6. Jeram S, Riego Sintes JM, Halder M, Baraibar Fentanes J, Sokull-Klüttgen B, Hutchinson TH. 2005. A strategy to reduce the use of fish in acute ecotoxicity testing of new chemical substances notified in the European Union. *Regul Toxicol Pharmacol* 42:218–224.
7. Organisation for Economic Co-operation and Development. 2009. The Threshold Approach for Acute Fish Toxicity Testing. Draft OECD Guidance Document. Paris, France.
8. European Commission. 2010. Directive 2010/63/EU of the European Parliament and the Council of 22 September 2010 on the protection of animals used for scientific purposes. *Official Journal of the European Union* L 276, 20.10. pp 33–79.
9. Belanger SE, Balon EK, Rawlings JM. 2010. Saltatory ontogeny of fishes and sensitive early life stages for ecotoxicology tests. *Aquat Toxicol* 97:88–95.
10. Weyers A, Sokull-Klüttgen B, Baraibar-Fentanes J, Vollmer G. 2000. Acute toxicity data: a comprehensive comparison of results of fish, *Daphnia* and algae tests with new substances notified in the EU. *Environ Toxicol Chem* 19:1931–1933.