Are single species toxicity tests alone adequate for estimating environmental hazard?

John Cairns, Jr.
Biology Department and University Center for Environmental Studies, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, U.S.A.

Keywords: single species toxicity tests, system level tests, application factors, hazard evaluation

Abstract

Most biologists agree that at each succeeding level of biological organization new properties appear that would not have been evident even by the most intense and careful examination of lower levels of organization. These levels might be crudely characterized as subcellular, cellular, organ, organism, population, multispecies, community, and ecosystem. The field of ecology developed because even the most meticulous study of single species could not accurately predict how several such species might interact competitively or in predator–prey interactions and the like. Moreover, interactions of biotic and abiotic materials at the level of organization called ecosystem are so complex that they could not be predicted from a detailed examination of isolated component parts. This preamble may seem platitudinous to most biologists who have heard this many times before. This makes it all the more remarkable that in the field of toxicity testing an assumption is made that responses at levels of biological organization above single species can be reliably predicted with single species toxicity tests. Unfortunately, this assumption is rarely explicitly stated and, therefore, often passes unchallenged. When the assumption is challenged, a response is that single species tests have been used for years and no adverse ecosystem or multispecies effects were noted. This could be because single species tests are overly protective when coupled with an enormous application factor or that such effects were simply not detected because there were no systematic, scientifically sound studies carried out to detect them. Probably both of these possibilities occur. However, the important factor is that no scientifically justifiable evidence exists to indicate that degree of reliability with which one may use single species tests to predict responses at higher levels of biological organization. One might speculate that the absence of such information is due to the paucity of reliable tests at higher levels of organization. This situation certainly exists but does not explain the lack of pressure to develop such tests. The most pressing need in the field of toxicity testing is not further perfection of single species tests, but rather the development of parallel tests at higher levels of organization. These need not be inordinately expensive, time consuming, or require any more skilled professionals than single species tests. Higher level tests merely require a different type of biological background. Theoretical ecologists have been notoriously reluctant to contribute to this effort, and, as a consequence, such tests must be developed by associations of professional biologists and other organizations with similar interests.

Introduction

Although this discussion may appear hostile to single species toxicity testing efforts, it is not intended to be. Single species tests are exceedingly useful and are presently the major and only reliable means of estimating probable damage from anthropogenic stress. However, many in the scientific community are certainly aware of the need for community and system level toxicity testing. How

Hydrobiologia 100, 47–57 (1983).

then does one account for the difference between awareness and performance? As an illustration of why such a difference exists, consider this scenario from a hypothetical workshop entitled 'The Contributions of Theoretical Ecology to Pollution Assessment'. On the first day of this workshop, reassuring exchange occurs among participants on ideas of energy flow, ecosystem dynamics, multiple aggregate variables, niche packing, and the like. On the second day, the group which must make use of this information confronts the theoretical ecologist with one or more site-specific problems and asks specifically how theoretical ecology can be used in a particular situation. It usually becomes abundantly clear that no system level measurements exist on which a concensus occurs among theoretical ecologists on use, interpretation, validity, and predictive value! Following this exercise, a retreat ensues to measurements associated with single species or at least those that are clearly not ecosystem level parameters. This is usually accompanied by a call for more research. A majority of scientists have attended such workshops; many have probably attended a number of such meetings with roughly similar scenarios.

The call for more research before specific recommendations can be made usually falls on deaf ears. This is a pity because truly more research is needed. I greatly fear that ecologists will lose credibility among practioners of pollution assessment because they have correctly called attention to rather vast and significant problems without following through with a professional concensus on which system level tests to carry out, measurements to make, methods to formally approve, and so on. There is even danger in calling attention to deficiencies in single species tests in predicting system level effects because it may cause some practioners and regulators to doubt the efficacy of any biological measurements. In fact, single species tests have proven remarkably effective to estimate responses at high levels of biological organization despite considerable theoretical deficiencies in using them. Nevertheless, if the field of environmental toxicology and chemistry is to continue to evolve, these deficiencies must be identified and corrective measures taken.

Another problem is the inescapable conclusion that research (to provide system level responses of the type just described) is not sufficiently theoretical for some funding agencies and too theoretical for others. There is some evidence that this problem has been recognized and has been addressed in a minor way. Other blocks to development of adequate ecosystem level tests include difficulties in getting specialists in necessary disciplines to collaborate when their salary increases and/or tenure and promotion may be judged by specialists with an uncharitable view toward group research.

A fundamental problem with present toxicity testing protocols is that they often estimate effects on an ecosystem as if the ecosystem were merely a collection of species exposed to a single pure compound under constant conditions. The need to go beyond single species testing to evaluate hazard to the environment posed by toxic chemicals is gaining momentum. A parallel thrust involving the study of increasingly complex systems for evaluating environmental fate of chemicals is also in progress. Although the outcome of these developments is not evident, it is abundantly clear that the need to examine both toxicity and environmental fate of chemicals in a more environmentally realistic way is now a *sine qua non.* As is the case with most developing fields, it would be unfortunate if these new approaches and new methods were used for regulation before a substantial and sound data base validating their efficacy has been produced. It would be equally foolish to retard development of such methods because they are not of immediate practical benefit. Both society and industry have much to gain from the production of more accurate means of estimating hazard of chemicals in the environment. While single species tests are far from perfect, their development has far outstripped development of determination of toxicological responses at higher levels of biological organization. Although this article addresses higher levels of biological organization than single species, it is not intended to denigrate the research at lower levels (e.g., enzyme) which might enable predictive capabilities for mechanisms of toxicity to be disclosed.

Robert MacArthur (1975) said 'Scientists are perennially aware that it is best not to trust theory until it is confirmed by evidence. It is equally true, as Eddington pointed out, that it is best not to put too much faith in facts until they have been confirmed by theory'. Since Hart *et al.* (1945) produced a method for toxicity testing that was soon endorsed by a committee of the Water Pollution Con-

trol Federation (Doudoroff *et al.* 1951), quite a large number of facts have been generated in the field of aquatic toxicology: water quality often may markedly mediate the expression of toxicity, some chemicals will interact synergistically or antagonistically, life history stages of a single species may not be comparably sensitive to different toxicants, and different species may alter their relationships to each other in terms of sensitivity to toxicants so that knowing response relationships to chemical A does not ensure prediction of relative sensitivity to chemical B. As a consequence, predictive value of toxicity tests has remained low, and transferability of information from one species to another and from one level of biological organization to another has not been satisfactory. Use of application factors to compensate for areas of ignorance or absence of data has not reached a stage of development where substantive scientific justification is available for the efficacy of the factors presently used. A certain degree of safety can be achieved by making the factors as large as the worse possible case demands, but sanitary engineers attempting to achieve these levels have found them technologically or economically impossible. In short, the period since 1945 might be characterized as an era where much evidence accumulated but little integrating theory surfaced. No integrating hypothesis was produced to pull these facts together or explain their relationship to each other.

Over the 30-year period from 1945 to 1975, an enormous toxicological data base for aquatic organisms had been generated. This data base was so large that it was beyond the capability of any individual to fully comprehend all details, even when the information was subdivided into a single group of organisms such as fish. However, an event occurred that showed that this very substantial data base was inadequate in several notable aspects: (1) the amount of information on a particular chemical was probably inadequate, (2) the kinds of information generated on a particular chemical were generally inadequate for making a scientifically justifiable estimate of hazard, and (3) the transfer of information on one chemical to estimate with precision the hazard of another did not appear feasible.

Under the provisions of the June 7, 1976, consent decree, the U.S. Environmental Protection Agency (USEPA) was directed to issue Federal Water Pollution Control Act effluent limitations and guidelines, new source standards of performance, and three treatment standards for 65 identified toxic pollutants. Implementation of the directives of this decree began when the USEPA drafted Water Quality Criterion Documents for each of the individual 65 pollutants. These criterion documents reviewed all pertinent literature available on the particular chemical and attempted to determine acceptable limits not to be exceeded for the protection of aquatic life and human drinking water supplies. The Water Quality Committee of the USEPA Science Advisory Board was charged with evaluating the 65 criterion documents. The report of this committee to the Science Advisory Board stated that no documents had conclusions that were scientifically justifiable. This report was accepted by the Executive Committee of the Science Advisory Board and transmitted to the Administrator of USEPA (then Douglas Costel). It is worth emphasizing that the court issuing the consent decree did not allow sufficient time for the USEPA to generate its own data base, and, therefore, USEPA was forced to prepare the criterion documents with data already available in the literature or documents in the open literature that generally were prepared for some other purpose. This event provided unmistakable proof that data to be used for the hazard evaluation must be systematically generated for that purpose.

Another event that had a major influence on toxicity testing and hazard evaluation was the passage of the Toxic Substances Control Act (TSCA) that became law with President Ford's signature on October 11, 1976. This act represents an attempt to establish a mechanism whereby the hazard of a chemical substance to human health and the environment can be assessed before the substance is introduced into the environment. The enactment of TSCA served as a powerful new stimulant to development of testing procedures to evaluate hazard associated with potentially toxic substances to human health and the environment.

These events and others clearly indicated the need for the development of a strategy for hazard evaluation that led to the production of a series of books now commonly referred to in the profession as Pellston I, II, III, and IV. Pellston I (Cairns *et al.* 1978) advocates linkage of the environmental concentration of a chemical (the term is meant to include such things as partitioning, transformation

processes, persistance, etc.) with the concentration producing no adverse biological effects. The degree of uncertainty or lack of confidence in estimating these two concentrations was a function of their proximity to each other. This view is summarized graphically in Fig. 1. Pellston II (Dickson *et al.* 1979) examines protocols used in various industrialized countries for systematically generating the data base necessary for a scientifically justifiable hazard evaluation. Pellston III (Maki *et al.* 1980) examines biotransformation processes and their role in estimating environmental concentration of a chemical. Pellston IV (Dickson *et al.* 1982) deals with modeling the fate of chemicals in the environment.

These and other publications had as a primary goal the development of an underlying strategy for estimating hazard. This strategy must be based on sound science and professional judgment and should be as cost-effective as possible. The most

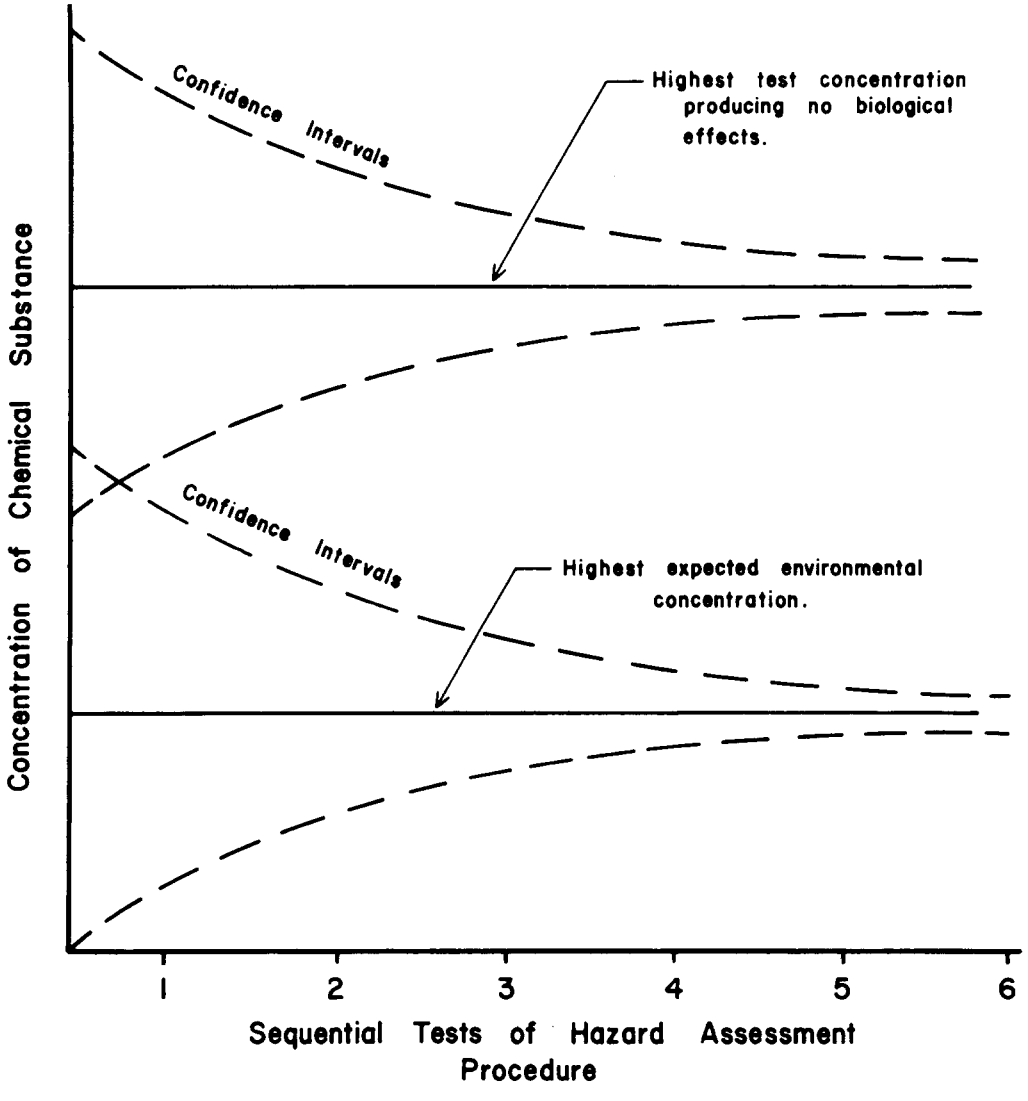

Fig. 1. Diagrammatic representation of a sequential hazard-assessment procedure demonstrating increasingly narrow confidence limits of estimates of no-biological effect concentration and actual-expected-environmental concentration. (Reprint with permission from ASTM STP 657, *Estimating the Hazard of Chemical Substances to Aquatic Life,* Copyright ASTM, 1916 Race St., Philadelphia, Pa., 19103.)

important consequence of these events just described has been to direct attention to the *information content* of data being generated (i.e., the facts) and *ways in which data will be used!* This will, in turn, add the additional requirement that data not only be precise, reliable, reproducible, and so on, but also *be suitable for the use or estimates proposed!* If the problem is viewed in this fashion, it becomes abundantly clear that the types of toxicity data now being generated are qualitatively deficient for their intended purpose! Toxicity tests should provide information that will facilitate predictions of the concentrations that will not harm living things in the environment *at all levels of biological organization!* The purpose of this manuscript is to present the view that single species tests alone are inadequate for this purpose.

Discussion

Some very important questions are related to testing at different levels of biological organization that deserve serious attention:

1. Can single species tests be used to predict responses reliably at other levels of biological organization?
2. In estimating the effects of chemicals on populations, multispecies assemblages, communities, and ecosystems, what are the limitations of laboratory science? In other words, are different degrees of environmental realism possible in the laboratory or under laboratory conditions at different levels of biological organization?
3. What should be the balance of toxicity tests at different levels of biological organization in order to make a valid estimate of hazard?
4. How should tests at different levels of biological organization be sequenced?
5. What criteria should be used to validate laboratory predictions in 'real world' field situations?

The complexity and uniqueness of each ecosystem has mitigated against ready transfer in general information from one site to another. Thus, many field studies are situation bound and highly site specific. Can the transferability of information from one site to another be enhanced by the development of mathematical models?

1. Can single species tests be used to predict responses reliably at other levels of biological organization?

One primary justification for using single species tests as a basis for estimating concentrations that will not prove harmful to communities and ecosystems is that if the most sensitive species is selected and concentration standards are set on that basis, then all other species will be protected. Since only a small percentage (probably less than 1%) of all freshwater species can be maintained in the laboratory sufficiently well to satisfy the requirement that no more than 10% of the control organisms expire during the course of tests, it seems quite unlikely that the most sensitive species will be selected for testing. In virtually every instance, the most sensitive species is being selected from a limited array of test species and extrapolating is being done from those results. Lest the discussion that follows be misunderstood, it is not intended to be an attack on single species toxicity testing. Such tests are essential for obtaining information on concentrations and durations of exposures to chemicals that result in changes in survival, reproduction, physiology, biochemistry, and behavior of individuals within particular species. One can question the scientific justification of using single species tests to predict changes in competition, predation, community function, ecosystem energy flow, and nutrient cycling. These are only a few of the many characteristics of ecosystems that either cannot be predicted from single species tests or for which there is insufficient evidence that the prediction is scientifically justifiable. Although practioners of single species toxicity testing may not state that the results of these tests can be used to protect biological systems of greater complexity than single species, the implication is definitely present. The public believes that when a concentration is purported to produce no adverse biological effects then the effects so described go beyond the kinds of effects that are characteristic of single species responses. If the field of environmental toxicology and chemistry is to prosper, 'truth in packaging' is mandatory in terms of the limitations, as well as the strengths, of single species toxicity tests now so widely used.

One common argument advanced by people who favor continuation of primary reliance on single species testing is that no significant ecological disas-

ters have occurred when carefully carried out single species tests were used. Of course, this could be merely due to the fact that single species tests are rarely validated by extensive, carefully carried out ecological investigations. It is not surprising that no adverse effects were noted because no extensive investigations were carried out to support this statement. In short, it is a statement based more on absence of information than on supporting information. It is quite likely that no dramatic events, such as a major fish kill, would be associated with waste discharge practices based on carefully carried out single species tests because, if the species were carefully selected and the test conducted by experienced professionals and a large application factor used, this should certainly not occur. On the other hand, changes in the ecosystem that might reduce fish population by impaired spawning rather than lethality would be less likely to be noticed by casual observers because no dramatic, highly visible evidence would be present to suggest major changes were occurring. It would be extremely helpful if predictions made with single species tests were validated by extensive field studies that would show whether or not both ecosystem structure and function were impaired at concentrations considered to have no adverse biological effects based on single species evidence alone.

There is also another intriguing possibility – single species tests are vastly overprotective. Ecologists have made statements for years that ecosystems are fragile because of their extraordinary complexity. The intuitively reasonable argument that such highly complex systems may be put into disequilibrium by disturbing any component of the system has been quite prevalent. The reasoning is that such an interdependent, interlooking system is fragile because of these abundant linkages. This complexity and multitudinous first, second, third, etc. order interactions are so well accepted by ecologists that any statement along these lines would be regarded as platitudinous. However, very little substantive evidence exists that supports the statement that complexity is necessarily associated with fragility. Some of the most complex ecosystems known to man are periodically subjected to major natural disturbances that they are either able to resist or, if displacement occurs, recover. In fact, Vogl (1980) and others have pointed out that some ecosystems actually deteriorate if striking disturbances do not

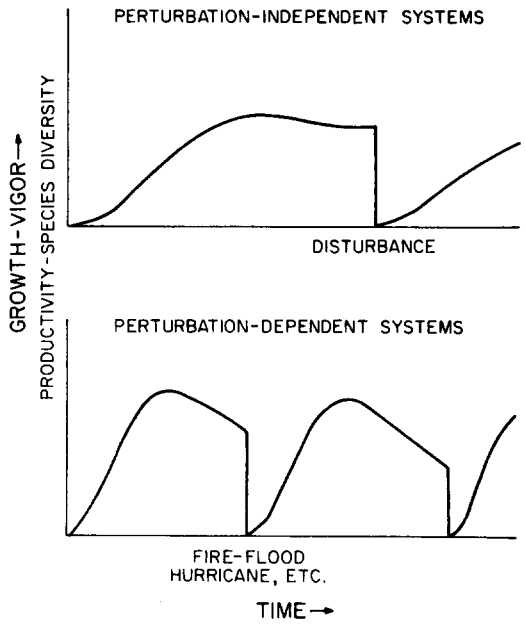

Fig. 2. Disturbances in general ecosystems create vegetational setbacks and complete recovery is slow, whereas disturbances in perturbation-dependent ecosystems usually stimulate pulses of growth which rapidly decline unless disturbed again.

occur at periodic intervals. The difference between disturbance-dependent and disturbance-independent ecosystems is given in Fig. 2. An alternative hypothesis equally tenable is that ecosystems are tough because they are complex and that damage to one of several similar pathways may result in shunting to alternative pathways of nearly comparable function. In this case, complexity would increase rather than decrease resistance to disturbance. Functional redundancy in ecosystems has been recognized for years. There may be several predators on a single prey species. The river continuum hypothesis (Vannote *et al.* 1980) indicates that certain processes, such as leaf degradation, may be carried out by different taxonomic groups in the upper and lower reaches of a stream. The end functional result, namely increased availability of nutrients and energy in the leaf, is unchanged.

In addition, the low environmental realism of the very simple, common toxicity test (where organisms are tested in a container with water but with no mud, rocks, vegetation, etc.) means that transformation of hazardous chemicals might occur less rapidly than in the 'real world'. Rapid transforma-

tion in the latter might produce secondary products less harmful than the original and result in decreased ecosystem vulnerability. Similarly, various types of environmental sinks for chemicals are not incorporated into most commonly used single species tests.

A final argument given by those who accept the need for going beyond single species testing will be that these are sufficient in instances when the estimated environmental concentration of the chemical is so far below the estimated no adverse effects concentration that it would be ridiculous to go beyond simple and inexpensive single species screening tests. Kimerle (1979) has noted, however, that the actual environmental concentration might be far higher than was estimated from simple laboratory screening tests and that the no adverse biological effects concentration might be far lower in the 'real world' than was estimated from simple screening laboratory tests (Fig. 3). In short, the screening tests did not accurately predict 'real world' events! In one case, the effects were vastly decreased and in the other (the environmental concentration) vastly increased (Fig. 3). The end result is two concentrations that appeared comfortably separated in the laboratory were in the 'real world' quite close together. Of course, the errors could be in the opposite direction in both cases and end up with two concentrations that appeared quite close together from laboratory evidence being quite dis-

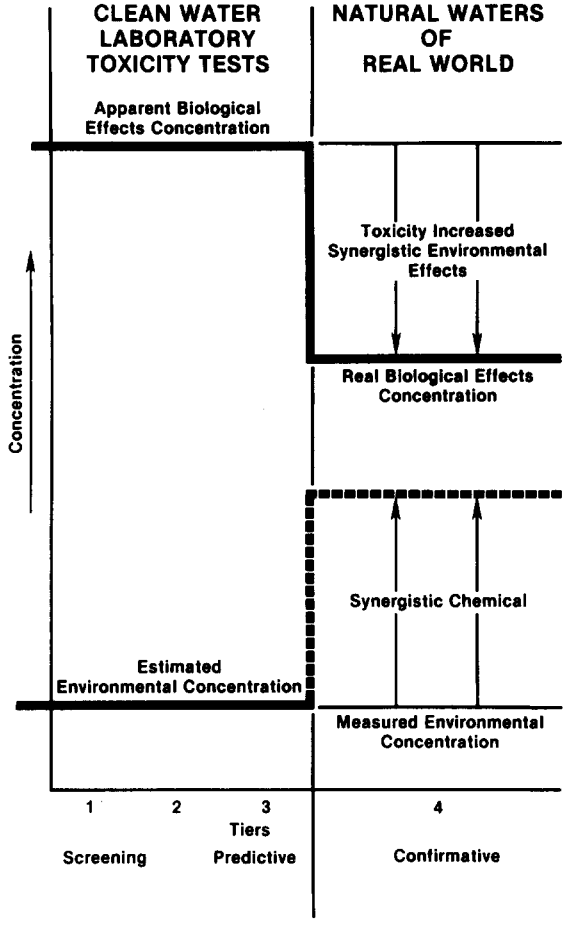

Fig. 3. Hypothetical situation of an apparent small margin of safety from clean water laboratory toxicity data actually being much smaller because of synergistic effects of natural waters. (From Kimerle, 1979, in *Workshop on Hazard Assessment.*)

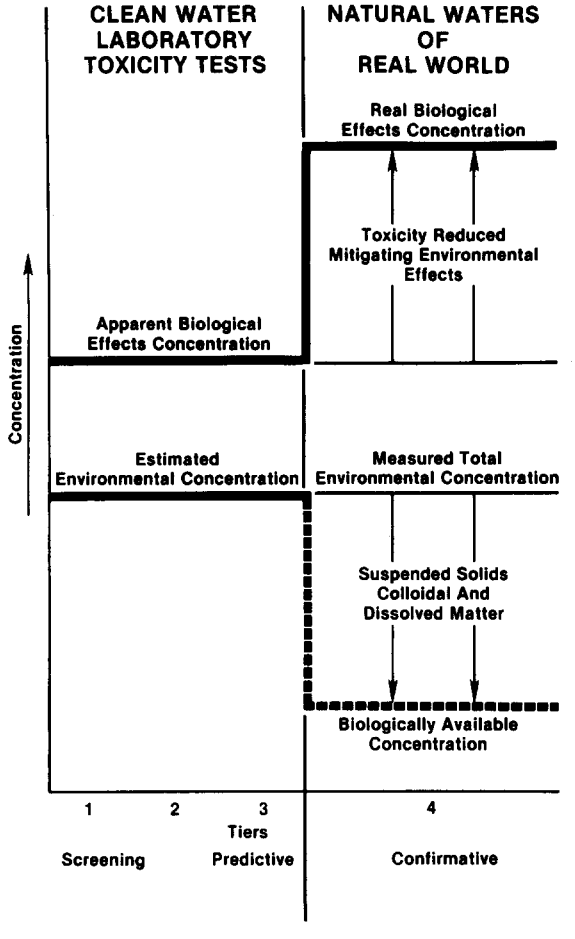

Fig. 4. Hypothetical situation of an apparent large margin of safety from clean water laboratory toxicity data actually being much greater because of mitigating effects of natural waters. (From Kimerle, 1979, in *Workshop on Hazard Assessment.*)

tant from each other in the real world (Fig. 4). These possibilities provide support for doing toxicity testing at more than one level of biological organization even for screening purposes.

From an economic standpoint, the soundest possible evidence on which to base management and regulatory decisions must be demanded. At the present time, sufficient evidence is not available to determine how accurately prediction can be done of toxicological response from one level of biological organization to another, but both theoretical biology and the rapidly accumulating data base on this subject seem to indicate that such predictions are relatively weak.

2. In estimating the effects of chemicals on populations, multispecies assemblages, communities, and ecosystems, what are the limitations of laboratory science?

Are there different degrees of environmental realism possible under laboratory conditions at different levels of biological organization? In a visiting scholar address given at the Mountain Lake Biological Station, Virginia, Eugene P. Odum (1981) gave an address entitled 'The Limitations of Laboratory Science' that beautifully illustrates some areas in which laboratory investigations, however skillfully carried out, will not suffice. It seems intuitively reasonable that environmental realism is more easily achieved in the laboratory at lower levels of biological organization (e.g., species) than at higher levels of organization (e.g., community or ecosystem). The report of the National Research Council Committee on Ecotoxicology (Cairns *et al.* 1981) takes note of the fact that situations occur where laboratory evidence will not be adequate and field testing will be mandatory. The report even makes a distinction between field situations in which the exposure is contained and those in which it is not. This merely recognizes that both effects of scale and time as well as degree of complexity must be considered that are not amenable to laboratory study. If this hypothesis is correct, then predictions of toxicological effects from one level of biological organization to another are not scientifically justifiable. In this event, an array of toxicity tests at different levels of biological organization are necessary for scientifically justifiable estimates of hazard. If the hypothesis is accepted that the limitations of laboratory science become greater as the complexity or level of biological organization increases, then both microcosms and field tests will become more common than they are now. Of course, both these hypotheses need careful and detailed testing so that the judgement of their soundness can be based on solid evidence. In addition, evidence must be obtained on the limitations of laboratory science in making predictions at different levels of biological organization. In other words, to what degree can more complex biological systems be simulated in the laboratory?

3. What should be the balance of toxicity tests at different levels of biological organization in order to make a valid estimate of hazard?

A rigid or specific balance of tests at each level of biological organization would not serve equally well for all situations and all categories of chemical substances. Therefore, protocols need to incorporate procedures for adjusting the balance of tests at different levels of biological organization as information on the level of complexity most likely to be affected is determined. Presumably, the most even distribution would be at the outset and become less and less even as critical sensitive components are identified.

The determination of balance, particularly as one proceeds through a sequential protocol at different levels of organization (i.e., a sequence for each of the major levels), will pose some interesting problems. For example, suppose that a particular single species test shows that deleterious effects would occur with that species but that major ecosystem functions would continue undisturbed. An additional condition, not explicitly stated but implied in the previous statement, could be made that alternative species were available to carry out similar ecological functions if the species suffering the deleterious effects represented a significant portion of the biomass. This would still pose a problem requiring considerable professional judgment and analysis because the loss of functional redundancy (i.e., reducing the number of species carrying out a particular function) is a deleterious ecosystem effect. An even more interesting decision would be forced by an effect on a transitory species that would soon be lost anyway because of normal successional processes with no other effects being discernible. A

number of other interesting arrays of mixed test results could be furnished that would illustrate the point that by adding more levels of biological organization to the test system that the increased complexity of the situation requires much more professional judgment than has ever been necessary. However, this merely emphasizes the point that judgments were probably being made on evidence that was far too simplistic.

4. How should tests at different levels of biological organization be sequenced?

Sequencing in a toxicity testing protocol can serve various purposes: (1) adding entirely different information from that already gathered, (2) expanding on information shown to be critical by earlier tests, (3) validating evidence gathered in previous tests. In many of the existing toxicity testing protocols, multispecies and system level tests are carried out only when the probability appears great that some deleterious effects might occur. In cases where the no adverse biological effects concentration (based on single species tests) was very markedly higher than the estimated environment concentration of the chemical, carrying out additional tests at higher levels of biological organization was usually considered unnecessary.

If several levels of biological organization are tested at the outset of a toxicity testing protocol, several alternative courses of action are possible: (1) if some or all the tests at various levels in the first part of the sequence show a probability of adverse biological effects or there is uncertainty about this probability, additional tests would be carried out at all levels of biological organization in the early part of the sequence unless compelling evidence exists for omission of one or more levels, (2) the most critical level of biological organization could be selected for these tests and additional tests could be conducted only in that sequence designed for that particular level (i.e., a sequence would be designed for each level of biological organization), (3) the levels of biological organization likely to give the least useful information could be eliminated but several levels each with its own sequence could be retained as a means of validating the presumed relationship identified in earlier parts of the sequence.

Since so few tests are now routinely utilized for hazard evaluation at levels of organization higher than the species, providing a detailed scenario on how they might be used is difficult. Some factors that would influence sequencing in the alternative system proposed would be the amount of information redundancy, the predictive capability within a particular level of organization from one function to another, and so on. Since so little information exists now on these factors, speculation on details of sequencing is difficult.

5. What criteria should be used to validate laboratory predictions in 'real world' field situations?

If predictions from one level of biological organization to another are unsound, validations of predications made using one level with a higher or lower level of biological organization would not work well. On the other hand, if a prediction is being validated, then it does not matter what level of biological organization was used in the test but only what level of biological organization is being protected by the prediction. Thus, in the first case the accuracy of a test in terms of the degree of environmental realism incorporated into the laboratory test is being validated. In the second case, the accuracy of the prediction based on the laboratory test is being validated. The prediction may be that ecosystem integrity will not be impaired at or below a certain concentration of a chemical. At the very least, putting this procedure into place would introduce a note of caution into the predictions being made, particularly where the protection of ecosystems is concernced. Furthermore, the basic assumptions underlying most present practices would be more rigorously examined. Finally, the ecologists would be forced to play a more active role as problem solvers and to endorse professionally those methods now available for making realistic ecosystem measurements and predictions. If none are immediately suitable for formal endorsement by professional ecologists, strenuous efforts would be made to determine why this situation exists.

Concluding remarks

When I entered the field of pollution assessment in 1948, the burning issue was whether biological testing had any role to play in pollution assessment.

Students in my classes find it difficult to believe that anyone would doubt the value of biological evidence, but an examination of the literature of the period from 1945 to 1950 will show this to be true (Patrick 1949). This was the period when the newly published simple batch toxicity testing method (Hart *et al.* 1945) was being acknowledged by a committee of the organization now called the Water Pollution Control Federation (Doudoroff *et al.* 1951) and was eventually incorporated as an American Society for Testing and Materials standard method. Had Hart *et al.* (1945) or the Doudoroff committee (1951) called for the various toxicity tests now commonly used with continuous flow requirements, embryo larvae tests, generational tests, tests at different trophic or functional levels, etc., they would have been regarded as hopelessly visionary. This is merely a consequence of the explosive and rapid development of a field that had only a few practioners in the late 1940s. However, the last 10 years have shown remarkable changes in both attitudes and methodology. Almost all the advances have occurred in single species toxicity testing. Testing at higher levels of biological organization has not kept pace with advances in single species testing, and an uncharitable person might say that a practioner using ecological methods of the late 1940s and early 1950s could still get by today.

There is abundant evidence, however, that a period of explosive development is already beginning in the use of laboratory microcosms, as well as in the use of artificial streams and larger simulation units which E. P. Odum has called *mesocosms.* Papers are beginning to appear in the professional literature validating laboratory tests in natural systems with a frequency that is in notable contrast to the virtual absence of such publications only a few years ago.

I recognize the considerable temerity of calling attention to the need for going beyond single species testing when such tests are just now being commonly used and no system level tests have been formally endorsed as standard methods. However, precise predictions of ecosystem effects will not be possible until methodologies and capabilities not now available have been developed. It has been said that looking into the future is equivalent to peering through a brick wall. Nevertheless, it seems intuitively reasonable that the following will be landmarks in the development of toxicity tests at higher levels of biological organization than single species:

1. The first professionally endorsed ecosystem level method appearing as a standard method in one of the presently recognized systems for doing so.
2. The first protocol in which system structure and function are given equal attention.
3. The first protocol in which tests at higher levels of biological organization than single species play a major role in generating the initial information on which subsequent testing is based.
4. The first formally endorsed field method for validating laboratory tests of any kind.

The November 1981 issue of *Science* announces as a matter of general interest the development of a sealed microcosm which has stable characteristics and species composition for over a year. Granted that this is a rather simple system, it nevertheless displays characteristics long sought by those who wish to carry out chronic microcosm tests under controlled conditions. Presumably now that this initial breakthrough has lead the way, additional methods will quickly appear as is often the case when a major new field begins to open.

Acknowledgements

This paper is based on invited presentations to the Society for Environmental Toxicology and Chemistry in November 1981 and to the Alberta Society of Professional Biologists in April 1982. I am indebted to Darla Donald for editorial assistance in the publication of the manuscript and to Judy Vest for typing it. Dr. M. A. Mehlman and Mobil Foundation supplied funds that made preparation of this manuscript possible.

References

Cairns, J. Jr., Alexander, M., Cummins, K. W., Edmondson, W. T., Goldman, C. R., Harte, J., Hartung, R., Isensee, A. R., Levins, R., McCormick, J. F., Peterle, T. J. & Zar, J. H., 1981. Testing for Effects of Chemicals on Ecosystems. Nat. Acad. Press, Wash. D.C. 103 pp.

Cairns, J. Jr., Dickson, K. L. & Maki, A. (eds.) 1978. Estimating the Hazard of Chemical Substances to Aquatic Life. Am. Soc. Test. Mater. Philad. Pa. 278 pp.

Dickson, K. L., Cairns, J. Jr. & Maki, A. (eds.) 1979. Analyzing the Hazard Evaluation Process. Am. Fish. Soc., Wash. D.C. 159 pp.

Dickson, K. L., Maki, A. W. & Cairns, J. Jr. (eds.) 1982. Modeling for Estimating the Fate of Chemicals in Aquatic Environments. Ann Arbor Sci. Publ., Inc., Ann Arbor, Mich., 413 pp.

Doudoroff, P., Anderson, B. G., Burdick, G. E., Galtsoff, P. S., Hart, W. B., Patrick, R., Strong, E. R., Surber, E. W. & Van Horn, W. M., 1951. Bioassay for the evaluation of acute toxicity of industrial wastes to fish. Sewage ind. Wastes 23: 1380–1397.

Hart, W. B., Doudoroff, P. & Greenback, J., 1945. The Evolution of the Toxicity of Industrial Wastes, Chemicals and Other Substances to Freshwater Fishes. Waste Cont. Lab., Atlantic Refining Co., Philad. Pa., 349 pp.

Kimerle, R. A., 1979. Aquatic hazard assessment – concepts and application. In: Bratzel, M. P., Jr. (ed.) Workshop on Hazard Assessment. Great Lakes Water Quality Board of the International Joint Commission, Windsor, Ontario, Canada: 221–230.

MacArthur, R., as cited in Cody, M. L. & Diamond, J. M. (eds.) 1975. Ecology and Evolution of Communities. Belknap Press of Harvard Univ. Press, Cambridge, Mass.

Maki, A. W., Dickson, K. L. & Cairns, J. Jr. (eds.) 1980. Biotransformation and Fate of Chemicals in the Aquatic Environment. Am. Soc. Microbiol., Wash. D.C. 150 pp.

Odum, E. P., 1981. Revival of non-laboratory science. As. S.-east. Biol. Bull. 28: 119–121.

Patrick, R., 1949. A proposed biological measure of stream conditions, based on a survey of the Conestoga Basin, Lancaster County, Pennsylvania. Proc. Acad. nat. Sci. Phila. 101: 277–341.

Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R. & Cushing, C. E., 1980. The river continuum concept. Can. J. Fish. aquat. Sci. 37: 130–137.

Vogl, R. J., 1980. The ecological factors that produce perturbation-dependent ecosystems. In: Cairns, Jr. Jr. (ed.). The Recovery Process in Damaged Ecosystems. Ann Arbor Sci. Publ., Inc., Ann Arbor, Mich.: 63–94.