onships, for example,
ınd fecundity of fish.
e, they can be used to
ɜ combined direct and
can potentially reveal
unreplicated, indirect
ıbility in such studies.
wetlands will result in
ıunity for any of their
consequences for all
ɜed ecosystem (Chap-
inties associated with

24 Testing

The toxicological literature is vast and largely trivial.

Moriarty (1988)

Toxicity tests are studies in which organisms, populations, or ecosystems are exposed to a chemical or mixture to determine the nature of effects and the relationship between the degree of exposure and effects. Similar tests may be performed for other agents. Examples include pathogenicity tests and tests of responses to physical/chemical conditions such as pH, dissolved oxygen, and suspended sediment. Tests resemble experiments in that the degree of exposure is controlled, exposures are replicated, assignment of replicate test subjects is randomized, and extraneous variance among replicates is minimized. They differ from classic scientific experiments in that they are intended to establish a functional exposure–response relationship. A test should determine not only that zinc can kill fish, or even that it kills them by disrupting ion exchange in the gills, but also how the proportion of fish killed increases with concentration or duration of exposure. This chapter describes tests of individual chemicals or materials, tests of contaminated media, and field tests.

The treatment of testing in this chapter is somewhat cursory, since the goal is to familiarize assessors with the types of test data being generated, not to teach them to perform tests. Detailed procedures are published by governments and standards organizations, which are cited in the appropriate sections. More detailed reviews can be found in ecotoxicology texts (Calow 1993; Rand 1995; Hoffman et al. 2003) or appropriate government documents such as Anderson et al. (2003).

24.1 TESTING ISSUES

Conventional toxicity tests determine effects on organisms and are divided into two classes, acute and chronic. Acute tests are those that last a small proportion of the life span of the organism (<10%) and involve a severe effect (usually death) on a substantial proportion of exposed organisms (conventionally 50%). Acute tests also usually involve well-developed organisms rather than eggs, larvae, or other early life stages. Chronic tests include much or all of the life cycle of the test species and include effects other than death (most often, growth and fecundity). Chronic test endpoints are typically based on statistical significance, so the proportion affected and the magnitude of affect may be large or small. In addition, there are many tests that fall between these two types that are termed subchronic, short-term chronic, etc. They typically have short durations but include sublethal responses. A prominent example is the 7 d fathead minnow test, which includes growth as well as death and includes only a small part of the life cycle, but that part is a portion of the larval stage (Norberg and Mount 1985).

In general, tests with longer durations, more life stages, and more responses are more useful for risk assessment, because they provide more information and because exposures in the real world are often sustained. However, if exposures are brief, acute or subchronic tests may be preferred. Examples include exposures of transients such as migratory waterfowl or highly mobile species that may use a site in transit or episodic exposures such as overflow of waste ponds, applications of pesticides, effluents generated during treatment failures, blow-down of cooling water, or flushing of contaminants into surface waters by storms.

The following are general recommendations for selecting tests of chemicals or materials. Other issues specific to tests of particular media or of other agents are addressed subsequently:

Standardization: In general, choose standard tests. Standard test protocols have been developed or recommended by governments (Keddy et al. 1995; EPA 1996b), and standard organizations (APHA 1999; OECD 2000; ASTM 2002). Most extrapolation models for relating test endpoints to assessment endpoints require standard data (Chapter 26). In addition, results of standard tests are likely to be reliable because methods are well developed, QA/QC procedures are defined, and test laboratories are likely to conduct standard tests routinely. However, nonstandard tests should be used when particular assessment-specific issues cannot be resolved by standard test results. Some effects such as effects of estrogenic chemicals on sexual development or the behavioral effects of lead are not observed in standard tests. Also, assessments may require tests of important local species or at least species that are relevant to a location. In particular, the biota of certain ecosystem types such as arid ecosystems are not well represented by standard tests (Markwiese et al. 2001).

Duration: Choose tests with appropriate durations. Two factors are relevant. The first is the duration of the exposures in the field. If exposures are episodic, as is often the case for aqueous contamination, tests should be chosen with durations as great as the longest episodes. The second factor is the kinetics of the chemical. Some chemicals such as chlorine in water or low-molecular-weight narcotics are taken up and cause death or immobilization in a matter of minutes or hours. Others have very slow kinetics and require months or years to cause some effects such as reproductive decrements. For example, tests of effects of polychlorinated biphenyls (PCBs) on mink have shown much greater effects on kit production and survival in the second year than in the first (Restum et al. 1998; Hornshaw et al. 1983).

Time course: Time to response is a neglected aspect of ecotoxicology and other components of applied ecology. In some cases, particularly episodic or accidental releases, variation in the duration of the exposure is more important to the risk estimate than concentration, which may be relatively constant or uncontrollable. In such cases, it is highly desirable to use data that report responses at multiple time points. For example, a report of a 96 h acute lethality test would be more useful if it reported the proportion surviving at each exposure level at 3, 6, 12, 24, 48, 72, and 96 h.

Response: Choose tests with appropriate responses. In particular, if an apparent effect of the contaminants has been observed in field studies, tests that include that effect as a measured response should be used. More generally, chosen tests should include responses that are required to estimate the assessment endpoint. The most common response parameters in toxicity tests are survival, fecundity, and growth, and most ecological effects models use one or more of these responses as parameters. Physiological and histological responses are generally not useful for estimating risks, because they cannot be related to effects at higher levels. However, if they are characteristic of particular contaminants, they can be useful for diagnosis (Chapter 4).

Media: Prefer tests conducted in media with physical and chemical properties similar to site media or media characteristic of the exposure scenario. For example, if assessing a pesticide for use on cotton, use tests in soils similar to those used for cotton production.

Organisms: Prefer taxa that are closely related taxonomically to the endpoint species and life stages that are likely to be exposed. If an assessment endpoint is defined in terms of a community, one may either choose tests of species that are closely related to members of the community or use all high-quality tests in the hope of representing the distribution of sensitivity in the endpoint community (Chapter 26). Species, life stages, and responses should also be chosen so that the rate of response is appropriate to the duration of exposure and kinetics of the chemical. In general, responses of small organisms such as zooplankters and larval fish are more rapid, because they achieve a toxic body burden more rapidly than larger organisms. Therefore, if exposures are brief and if those small organisms are relevant to the assessment endpoint, tests of small organisms should be preferred over larger organisms that are no more relevant. However, such tests may not be appropriate if, for example, the endpoint is fish kills. Finally, choose taxa and life stages that are known to be sensitive to the agent being assessed.

Multiple exposure levels: Studies that employ only a single concentration or dose level plus a control are seldom useful. If the exposure causes no effect, it may be considered a no observed effects level (NOEL), but no information is obtained about levels at which effects occur. If the exposure causes a significant effect, it may indicate that a reduction in exposure is required, but the necessary magnitude of reduction cannot be determined. Studies in which multiple exposure levels were applied allow an exposure–response relationship to be evaluated and thresholds for effects to be determined. Consequently, studies that applied multiple exposure levels are strongly preferred.

Exposure quantification: To correctly interpret the results of toxicity tests and to apply these results in risk assessments, the exposure concentrations or doses should be clearly quantified. Ideally, the test chemical should be measured at each exposure level; measured concentrations are preferable to nominal concentrations.

Chemical form: Correct estimation of exposure requires that the form of toxicant used in the test be clearly described. For example, in tests of lead, the description of the dosing protocol should specify whether the dose is expressed in terms of the element (e.g., lead) or the applied compound (e.g., lead acetate). Tests of chemicals in the forms occurring on the site are preferred. This is particularly important for chemicals that may occur under ambient conditions in multiple ionization states or other variant forms that have differing toxicities.

Statistical expressions of results: The traditional toxicity test endpoints for chronic tests, NOELs and lowest observed effect levels (LOELs), have been used to establish benchmarks or criteria (Chapter 29), but they have low utility for risk assessment, because they are based on statistical significance rather than biological significance (Chapter 23). To fully estimate risks, it is necessary to estimate the nature and magnitude of effects that are occurring or could occur at the estimated exposure levels. To do this, exposure–response relationships should be developed for chemicals evaluated in ecological risk assessments.

These criteria may conflict in some cases, because the best test data for one criterion may not be the best for another. Therefore, assessors must judge their relative importance to the particular assessment, and apply them accordingly.

24.2 CHEMICAL OR MATERIAL TESTS

In ecological risk assessments, effects data for single chemicals, organisms (e.g., an exotic parasitoid), or materials (e.g., gasoline, silt) may be obtained from tests performed ad hoc (primary data), but are usually obtained from the literature or from databases (secondary and tertiary data). One useful tertiary source is the EPA ECOTOX database (http://www.epa.gov/medatwrk/databases.html), which contains toxicity data for aquatic biota, wildlife, and terrestrial plants. Reviews such as those produced by R. Eisler for the US National

Biological Service are also useful tertiary sources (www.pwrc.usgs.gov/new/chrback.htm). Test data generated for an assessment (primary data) are relevant by design, but when data are taken from the literature or from reviews (secondary or tertiary data), assessors must select those data that are most relevant to the assessment endpoints and that can be used with the exposure estimates, as discussed in the previous chapter. However, because the variance among chemicals is greater than the variance among species and life stages, any toxicity information concerning the chemicals of interest is potentially useful. If no toxicity data are available that can be applied to the assessment endpoints (e.g., no data for fish or no reproductive effects data), or if the test results are not applicable to the site because of differences in media characteristics (e.g., pH or water hardness), tests may be conducted ad hoc. If combined toxic effects of multiple contaminants are thought to be significant, and if appropriate mixtures are not available in currently contaminated media, synthetic mixtures may be created and tested, or combined effects models may be applied (Chapter 8).

Test data from the literature have biases that should be understood by ecological risk assessors. Assessors must be aware of these biases when test data are used to derive toxicity benchmarks or exposure–response models for chemicals. Potential sources of bias in test data include:

- *Form*: The forms of chemicals used in toxicity tests are likely to be more toxic than the dominant forms in the field. For metals the tested forms are usually soluble salts, and organic chemicals may be kept in aqueous solution by solvents. In oral dosing tests, organic chemicals are often dissolved in readily digested oils.
- *Species*: The test species may not be representative of the sensitivity of species native to the site.
- *Media*: The standard media used in toxicity tests may not be representative of those at a particular contaminated site. For example, aqueous tests typically use water with moderate pH and hardness with little suspended or dissolved matter, and soil tests typically use agricultural loam soils or similar artificial soils.
- *Conditions*: Laboratory test conditions are less variable and may not be representative of field conditions (e.g., optimum temperature, sieved soil, or constant moisture).

24.2.1 Aquatic Tests

More test data are available for aquatic biota than any other type of ecological receptors (Table 24.1). In general, flow-through tests, which constantly renew the test water, are preferred over static-renewal tests, which renew the water periodically, and those in turn are preferred over static tests, which do not change the water. Flow-through tests maintain constant concentrations, whereas concentrations may decline significantly in static tests or even static-renewal tests due to evaporation, degradation, sorption, etc. However, static tests may be appropriate for extremely short-duration tests. The most abundant type of test endpoint is the 48 or 96 h median lethal concentration (LC_{50}). Life cycle tests that include survival, development, and reproduction provide the most generally useful data, but they are largely restricted to short-generation invertebrates because of the expense of long tests. For fish, early life stage tests of survival and growth are most commonly used, based on the presumption that early life stages are the most sensitive (McKim 1985). However, reproduction is often the most sensitive life stage (Suter et al. 1987), and, even if embryos or larvae are the most sensitive stage, maternal transfer may be an important route of exposure. These concerns may be addressed by short-term fish reproduction tests (Ankley et al. 2001), but for some chemicals, only a full life cycle test will reveal the effects on reproduction of long-term exposures of the adult fish.

TABLE 24.1
Examples of Standard Aquatic Toxicity Tests Published by the US EPA or the American Society for Testing and Materials (ASTM)

Taxon	Type	Reference[a]
Fish	96 h LC_{50} (juvenile or adult)	EPA/660/3-75-009 EPA/600/4-90/027F EPA/712-C-96-118 ASTM E729-96, -88
Fish	Early life stage survival and growth (egg through juvenile)	ASTM E 1241-97 EPA/712-C-96-121
Fish	7 d larval survival and growth	EPA/600/4-91/002 EPA/600/4-95/136 EPA/600/4-91/003
Macroinvertebrates	48–96 h LC_{50}	EPA/660/3-75-009 ASTM E729-96, -88
Mysid (saltwater crustacean)	Life cycle test	EPA/712-C-96-166 ASTM E 1191-97
Daphnia	Life cycle test	EPA/712-C-96-120 ASTM E 1193-97
Ceriodaphnia	7 d survival and reproduction	EPA/600/4-91/002
Algae	96 h growth	EPA/712-C-96-164 ASTM E 1218-97a

[a]EPA method reports may be obtained by searching www.epa.gov for the listed report number.
ASTM methods may be purchased by standard number from www.astm.org.

Dietary exposures may be important contributors to toxicity for bioaccumulative organic chemicals and metals, but are seldom tested. This is in part because of the difficulty of culturing or collecting contaminated food organisms or of realistically contaminating artificial diets. It also reflects a lack of general acceptance of the importance of aquatic dietary exposures. Hence, most aquatic dietary toxicity studies have been concerned with demonstrating the reality and nature of the problem rather than generating relevant exposure–response relationships (Meyer et al. 2005).

Currently, the most popular freshwater test organisms in the United States are fathead minnows (*Pimephales promelas*) and daphnids (*Daphnia* spp. and *Ceriodaphnia dubia*). The most common saltwater organisms are sheepshead minnows (*Cyprinodon variegatus*) and the mysid shrimp (*Americamysis bahia*). Test results for algae (often *Selinastrum capricornutum*) and aquatic plants (often duckweed, *Lemna gibba*) are less abundant than for aquatic animals. These tests have short durations (72 to 96 h), but they include multiple generations of vegetative reproduction. Further, plant tests usually report growth (e.g., cell or frond number) or production (e.g., carbon fixation rates), which are often applicable to assessment of risks to ecosystems.

24.2.2 Sediment Tests

Selecting representative sediment tests and test results is complicated by the interactions among the multiple phases (i.e., particles, pore water, and overlying water) of the sediment system. Sediment tests may be conducted using whole sediment or aquatic tests may be used to represent one of the aqueous phases. Test selection depends on the expected mode of

exposure, and more than one test type may be appropriate. Spiked sediment tests consist of the addition of known quantities of the test chemical or material to a natural or synthetic sediment to which the test organism is exposed. Spiked sediment tests provide an estimate of effects based on all direct modes of exposure, including ingestion, respiration, and absorption. Hence, toxicity to sediment ingesting organisms may be best approximated by spiked sediment tests. The primary disadvantage is the uncertain applicability of the exposure–response results to any particular field sediment or even to the distribution of field sediments. Aqueous tests are most appropriate if interstitial or overlying water is believed to be the primary exposure pathway for the toxicants and receptors at a site.

Aqueous tests are much more common than spiked sediment tests, but few aqueous tests use benthic species. Conventional aqueous tests and data are used to evaluate aqueous-phase exposures of benthic species, based on data suggesting that benthic species are not systematically more or less sensitive than water column species (EPA 1993d). For nonionic organic chemicals, aqueous concentrations and sediment concentrations can be interconverted by assuming equilibrium partitioning between the aqueous phases and the organic matter in the solid phase (Section 22.3).

An adjustment is also available for some sediment metals, based on the acid volatile sulfide (AVS) component of sediment (Section 22.3). However, it does not serve to estimate aqueous concentrations or effects.

When spiked sediment tests are used, physical and chemical properties of the test media are particularly important for evaluating chemical toxicity. Characteristics of the sediment (e.g., organic carbon content and grain size distribution) and of water (e.g., dissolved organic carbon, hardness, and pH) can significantly alter the speciation and bioavailability of the tested material. In site assessments, tests in sediments similar to the site media should be preferred. Regression models could be derived to account for confounding matrix factors (e.g., grain size or organic carbon content) (Lamberson et al. 1992). However, such models are species- and matrix factor-specific and would need to be developed on a case-by-case basis. The test method also can affect exposure. For example, chemical concentrations and bioavailability can be altered by the overlying water turnover rate, the water/sediment ratio, and the oxygenation of the overlying water (Ginn and Pastorok 1992). For all of these reasons, spiked sediment tests are relatively uncommon.

24.2.3 Soil Tests

There are relatively few standard soil tests, and the body of published toxicity data is small relative to water and sediment tests. In particular, few organic chemicals other than pesticides are represented. Standard methods using spiked soil or solutions are available for vascular plants (mainly crops) and earthworms from the US Environmental Protection Agency (US EPA) (OPPTS 850 test guidelines), American Society for Testing and Materials (ASTM; Committee E47 standards), European Union, and others. A variety of additional tests have been developed, mostly in Europe (Donkin and Dusenbery 1993; Donker et al. 1994; van Gestel and Van Straalen 1994; Kammenga et al. 1996; Heiger-Bernays et al. 1997; Lokke and van Gestel 1998). These tests generally treat soil as a medium in which particular species are exposed, rather than as an ecosystem. Tests of effects on field soils with associated communities are described in Section 24.5.3.

Tests in both spiked soil and aqueous solutions may be useful for assessing risks from soil contaminants. The relevance of published tests in soil to the assessment of risks to soil organisms seems self-evident, but because soil properties are highly variable and greatly influence toxicity, the toxicity in any other soil may be quite different. For example, Zelles et al. (1986) found effects of chemicals on microbial processes to be highly dependent on soil

type. Unrealistic extremes of that variance should be eliminated by excluding data from tests in quartz sand, peat, or vermiculite, unless toxicity of chemicals mixed with these materials is demonstrated to be similar to that in natural soils. Tests conducted in solution have potentially more consistent results than those conducted in soil. Toxicity observed in inorganic salt solution may be related to concentrations in soil extracts, estimated pore water concentrations, or contaminated springs in which wetland plant communities are located. It has even been proposed that aquatic toxicity test results could be used to estimate the effects of exposure of plants and animals to contaminants in soil solution (van de Meent and Toet 1992; Lokke 1994), but that practice is not generally accepted.

Test endpoints for soil tests are less standardized than those for aquatic or wildlife tests. Plant tests most commonly include germination or growth. Invertebrate tests most commonly include survival but sometimes include reproduction. Tests of litter-feeding earthworms may not be representative of those that ingest soil, and vice versa. Similarly, although pollution-induced community tolerance (PICT) (Rutgers et al. 1998) has been used as a toxicological endpoint, it is not always clear that microbial communities that have become altered in their tolerance of contaminants are indicative of a decrease in the rate of a valued microbial process (Efroymson and Suter 1999).

24.2.4 Oral and Other Wildlife Exposures

Terrestrial and semiaquatic wildlife are exposed through oral, dermal, and inhalation routes, and by intergenerational transfers. Tests exist for each of these routes, but oral tests to estimate the effects of toxicants in food, water, or other orally consumed materials are most common. These tests are employed primarily with birds and mammals.

For dietary tests, test animals are allowed to consume food or water ad libitum that has been spiked with the test material. The amount of food consumed should be recorded daily so that the daily dose can be estimated. A potential problem with dietary tests is that animals may not experience consistent exposure throughout the course of the study. For example, as animals become sick (e.g., due to toxicity), they are likely to consume less food and water. They may also eat less or refuse to eat if the toxicant imparts an unpleasant taste to the food or water or if the toxic effects induce aversion. These problems are sufficiently serious for cholinesterase-inhibiting pesticides and some other chemicals that the use of dietary tests is questionable (Mineau et al. 1994).

In oral tests, animals receive periodic (usually daily) toxicant doses by gavage (i.e., esophageal or stomach tube) or by capsules. The chemical is usually mixed with a carrier such as water, mineral oil, or acetone solution to facilitate dosing. Oral tests may include a single dose to simulate an isolated and brief exposure, or daily doses to simulate a continuous or long-term exposure. They provide better-defined exposures and can be representative of oral exposures other than food or water consumption such as incidental soil ingestion or oral ingestion of oil or other materials during grooming.

The choice of carrier used for oral or dietary tests has been shown to influence uptake by binding with the toxicant or otherwise influencing its absorption. For example, Stavric and Klassen (1994) reported that the uptake of benzo(a)pyrene by rats is reduced by food or water but facilitated by vegetable oil. Similarly, uptake of inorganic chemicals varies dramatically between tests with food and water. Chemicals are generally taken up more readily from water than from food.

Results of most dietary toxicity tests are presented as toxicant concentrations (mg/kg) in food or water. These data can then be converted into doses (mg agent/kg body weight/d) by multiplying the concentrations in food or water by food ingestion rates and dividing by body weights either reported in the literature or presented in the study (Section 22.8).

Standard methods for avian acute, subacute, and reproductive oral toxicity tests have been developed (Table 24.2). In general, risks to mammalian wildlife are assessed using the same tests of laboratory rats and mice that are used in human health risk assessments, but wild mammal tests may be required when the particular issues suggest that those tests may not be adequate.

Dermal and inhalation test results may be found for rodents and sometimes other species in the mammalian toxicity literature. The methods developed for laboratory test species may be adapted for mammalian wildlife (see US EPA guidelines OPPTS 870.3465 and .3250). Methods for testing dermal or inhalation exposures for birds, reptiles, or amphibians must be developed largely as needed.

Effects of developmental toxicants on birds and other oviparous species are readily tested by egg injection. For example, effects of dioxin-like compounds, for which embryo development is critically sensitive, have been tested by injecting the eggs of chickens and other species (Hoffman et al. 1998).

24.3 MICROCOSMS AND MESOCOSMS

Microcosms and mesocosms, together termed model ecosystems, are physical representations of ecosystems that contain multiple species and usually multiple media and that may be replicated. Microcosms are small enough to be maintained in the laboratory. They include everything from a mixed microbial culture in a beaker to aquaria and small artificial streams. Mesocosms are larger, more complex, and located out of doors. They include artificial streams and ponds and enclosed areas of terrestrial, wetland, and shoreline ecosystems. Microcosms and mesocosms have similar purposes and overlap in terms of size and complexity. In general, the purposes for conducting tests in these systems are to (1) provide realistic fate and exposure kinetics by including degradation, sorption, and uptake; (2) include all routes of exposure; (3) expose a large number and variety of types of organisms; (4) allow secondary effects due to species interactions; and (5) allow the operation of ecosystem processes. The biotic components of microcosms may be specified so as to improve replication and understanding of responses (Taub 1969, 1997). More commonly, the microbial and invertebrate components of these systems are provided by the process of enclosure or by inoculation from a natural ecosystem. For example, a pond microcosm or mesocosm might be inoculated with sediment and water from a pond or lake that contains microbes and invertebrates. Prescribed numbers of fish, amphibians, or other large organisms may then be added to the inoculated mesocosm or large microcosm.

The use of model systems in the assessment and regulation of chemicals has been a source of ongoing controversy. Advocates argue that if the goal is to protect ecosystems, one must test ecosystems (Cairns 1983), but they have disagreed about the designs. Opponents argue that these simplified systems do not capture critical properties of real ecosystems (Carpenter 1996; Schindler 1998). Even advocates recognize that different ecosystems respond in qualitatively different ways, which are only sometimes reflected in model systems (Harrass and Taub 1985). Hence, after decades of development and advocacy, these test systems are still rarely used and have had relatively little influence on environmental decision making. Many issues have contributed to the controversy, some of which are discussed here.

Ecosystem definition: Microcosms or mesocosms have been said to represent ecosystems in the sense that a rat represents mammals and a fathead minnow represents fish. That is, ecosystem responses like changes in net production or species number are thought to be reasonably consistent across flasks, ponds, and rivers. On the other hand, ecosystems are

TABLE 24.2
Selected Standard Oral Toxicity Methods for Birds

Test Type	Test Species	Duration	Exposure Route	Test Endpoint(s)	Reference[a]
Acute	Northern bobwhite, Mallard	14 d	Gavage	Mortality, intoxication	OPPTS 850.2100
Subacute dietary	Northern bobwhite, Japanese quail, Mallard, Ring-necked pheasant	5 d exposure, 3 d post	In diet	Mortality, intoxication, other effects	OPPTS 850.2200 ASTM E857-87
Reproduction	Northern bobwhite, Mallard	10 week	In diet	Adult mortality Eggs laid Egg fertility Egg hatchability Eggshell thickness Weight and survival of young	OPPTS 850.2300

[a]EPA methods reports may be obtained by searching www.epa.gov for the listed report number. ASTM methods may be purchased by standard number from www.astm.org.

much less consistent in terms of their components and function than classes of organisms, and might not be expected to respond as consistently to a particular exposure.

Size: Model systems vary greatly in size. Smaller systems provide more replication at less cost, but larger systems can support more types of organisms and perhaps better represent ecosystems of concern.

Composition: No model system can support the full range of taxa and trophic levels of a real lake, river, or forest. However, it is not clear how much simplification is allowable before a microcosm or mesocosm fails to be a model of ecosystems of concern.

Type of response: Are model systems best used to realistically expose organisms, elucidate species interactions, measure ecosystem properties, document recovery, or something else?

Model systems vs. system models: Mathematical models of ecosystems (Chapter 28) and physical models (microcosms and mesocosms) potentially serve the same purpose in ecological risk assessment. While physical models are clearly more realistic, mathematical models are more flexible in terms of being able to represent a variety of systems, states of the systems, and conditions of exposure, and they are much less expensive to implement. Physical models require a major assumption; they represent the real ecosystems of concern. Mathematical models require an equivalent major assumption (e.g., that the response of the system can be represented by a trophodynamic model) plus many assumptions associated with their equation forms and parameter values.

Most of these issues are a result of the difficulty in defining ecosystems compared to organisms. When testing fishes or birds, one does not need to decide how many livers or eyes to include, how big the tested piece should be (use the whole organism), or even what the basic responses are (there is a consensus for survival, growth, and reproduction). This is because organisms are clearly defined entities, not just representatives of a level of organization. Hence, assessors must define assessment communities but not assessment organisms (Box 16.6). While some of the problems have been addressed by developing standard designs, these fundamental issues have led to problems of interpretation that have severely limited the utility of these systems for regulation or management. In particular, the US EPA developed a standard aquatic mesocosm for tier IV testing of pesticides (Tuart 1988), but dropped the requirement for these tests in 1992. This decision was based on the judgment that field tests had not significantly improved the bases for most registration decisions over laboratory tests, in part because of the ambiguities in interpreting their complex and highly variable results. Also, it was judged that post-registration field monitoring could serve the need for field results in most cases (Tuart and Maciorowski 1997). Problems in interpreting results of microcosms and mesocosms were addressed by a recent workshop (Giddings et al. 2002). The recommendations include the following points that are particularly relevant to assessors:

Design: Tests should be designed to provide exposure–response relationships, not just to test a particular predicted exposure level.

Types of endpoints: Structural and functional endpoints are generally equally important, but species structure is primary, and functional endpoints alone do not protect biodiversity.

Recovery: Initial effects should not be considered unacceptable if population recovery occurs in an acceptable time, and they do not cause adverse indirect effects.

Models: Ecological models for extrapolation to real ecosystems or to other exposures should be developed.

Data for extrapolation: Biological and physical conditions of the actual ecosystems to be simulated by the model systems should be determined to aid extrapolation.

Scenarios: The agricultural landscape or other landscape context should be used to design reasonable exposure scenarios.

Endpoints: Regulatory authorities must develop goals and assessment endpoints that test systems can be designed to support.

Training: Tests in model systems are difficult to conduct and their responses are complex and difficult to interpret, so guidance, training and tools are required.

An important additional recommendation is that users of model systems should state their assumptions (Clements and Newman 2002). Like mathematical models, physical models are simplifications of real systems, and those simplifications imply assumptions about what is and is not important to understanding the response of the real systems being simulated. For example, an aquatic microcosm may require the assumptions that planktivorous fish are not important to understanding effects on plankton and that macrophytes are dominant components of the ecosystems being assessed.

Aquatic model systems range from flasks in the laboratory to artificial ponds and streams (Graney et al. 1994, 1995; Kennedy et al. 2003). The following are major types of model ecosystems:

Standard aquatic microcosms: This system is assembled in a flask from sterile sand and water, ten species of algae, five species of zooplankters, and a bacterium plus inadvertently introduced microbes (Taub and Read 1982; OPPTS 1996a; Taub 1997).

Pond microcosms: Pond water and sediment, with associated microbes and invertebrates, are placed in flasks, aquaria, or tanks, sometimes with macrophytes but seldom with fish (Giddings 1986).

Pond mesocosms: Ponds ranging from 0.04 to 0.1 ha are dug and filled with sediment and water from a real pond or lake. Macrophytes and fish may then be added. These systems have often been used to study the fate and effects of pesticides in the United States and Europe (Tuart and Maciorowski 1997; Campbell et al. 2003).

Artificial streams: Indoor or outdoor channels may be treated with once-through or recycling water. Unlike pond mesocosms these systems have not been standardized and they range from small channels that support algae and invertebrates to in-ground channels with pools and riffles that are large enough to support fish (Graney et al. 1989).

Littoral enclosures: Replicate systems are created by enclosing 50 m^2 portions of a pond or lake and adjoining shoreline (Brazner et al. 1989; Lozano et al. 2003).

Limnocorrals: Portions of a littoral ecosystem are enclosed by a large plastic bag or cylinder suspended from a floating platform and anchored to the bottom. They vary in volume from 1000 to 100,000 L (Graney et al. 1995).

Microcosm tests of the soil community and processes such as decomposition incorporate indirect effects of chemical addition as well as direct toxic effects. In the United States, standard test of soil function determines effects on respiration, ammonification, and nitrification in sieved soil (Suter and Sharples 1984; OPPTS 1996f), or nutrient retention, respiration, and plant production in soil cores (Van Voris et al. 1985; OPPTS 1996b). Other soil microcosms are used to test effects on soil community structure (Parmelee et al. 1997).

Soil microcosm tests usually focus on changes in the rates of soil microbial processes, which, however, may increase or decrease in response to a chemical exposure. Some metals are nutrients at low concentrations and most organic chemicals are microbial substrates. For example, the antibiotic streptomycin reduces bacterial abundance but increases overall soil activity by serving as a carbon and nitrogen source (Suter and Sharples 1984). Further, reductions in some soil processes such as litter decomposition may be desirable or acceptable in particular ecosystems (Efroymson and Suter 1999). A litter layer is esthetically desirable, reduces erosion, and is important to successful germination of some trees, but introduced earthworm species are destroying litter layers in forests of the northeastern United States. Therefore, if soil processes are assessment endpoints, it is desirable to determine the relevant exposure–response relationship and to understand the ecosystem context of the processes.

Mesocosm studies of wildlife are much less common. Even more than aquatic mesocosms, they are primarily used to study effects of realistic exposures, secondarily to reveal secondary effects, and very little to show population-level effects. For example, Dieter et al. (1995) placed mallard ducklings in littoral mesocosms to evaluate effects of aerial application of the organophosphate insecticide phorate on waterfowl in prairie wetlands. In another study, Barrett (1968) evaluated the effects of the carbamate insecticide carbaryl on plants, arthropods, and small mammals in 1 acre old-field enclosures.

24.4 EFFLUENT TESTS

Standard toxicity tests have been developed for determining the acceptability of aqueous effluents and are widely used in effluent permitting in the United States. Although conventional acute lethality tests may be used, short-term chronic tests using short-lived species or subchronic tests using sensitive life stages are used (Table 24.3). These tests are unique in the extent to which they have been validated against biosurvey data (Mount et al. 1984; Birge et al. 1986; Norberg-King and Mount 1986; Dickson et al. 1992, 1996). In those studies, the 7 d fathead minnow and *C. dubia* tests have been found to be predictive of reductions in the species richness of aquatic communities. As a result of this intensive development and validation, these tests are widely used beyond the regulation of aqueous effluents, and many laboratories are available to conduct them. Other species and taxa are used outside the United States (Herkovits et al. 1996). Effluent tests may pass or fail, i.e., only the undiluted effluent or only one critical dilution may be tested to determine acceptability. However, it is preferable to test a series of dilutions to establish the exposure–response relationship (EPA 2002b). Most effluent tests are static-renewal, but static tests may be used if the effluent has little oxygen demand and the toxic constituents are not lost through volatilization or other processes. If tests are performed on site, flow-through effluent tests are possible.

Effluent tests may also be used to identify which components of the contaminant mixture are responsible for effects, a process called Toxicity Identification Evaluation (TIE) (EPA

TABLE 24.3
Standard Procedures Used to Test the Toxicity of Effluents and Ambient Waters[a]

Species	Life Stage	Response	Duration (d)	Medium[b]
Sea urchin	Eggs and sperm	Fertilization	0.3	SW
Daphnia sp.	Juvenile	Immobilization	2	FW
Bivalve mollusk	Larvae	Mortality, shell development	2	SW
Fish[c]	Juvenile	Mortality	4	FW/SW
Algae (*Selenastrum capricornutum*)	Cell culture	Growth	4	FW
Ceriodaphnia dubia or *Mysidopsis bahia*	Juveniles–adults	Immobilization, fecundity	7	FW/SW
Fish[c]	Larvae	Mortality, growth	7	FW/SW
Fish[c]	Embryo–larvae	Mortality, deformities	7	FW/SW
Algae (*Champia parvula*)	Culture	Sexual fecundity	7	SW

[a]Test protocols are found in (EPA 2002b,h,i) and ASTM standards.
[b]FW = freshwater; SW = saltwater; FW/SW = protocols are available for species from both media.
[c]The standard freshwater fish in the United States is the fathead minnow (*Pimephales promelas*) and the saltwater fish is the sheepshead minnow (*Cyprinidon variegatus*) or inland silverside (*Menidia beryllina*).

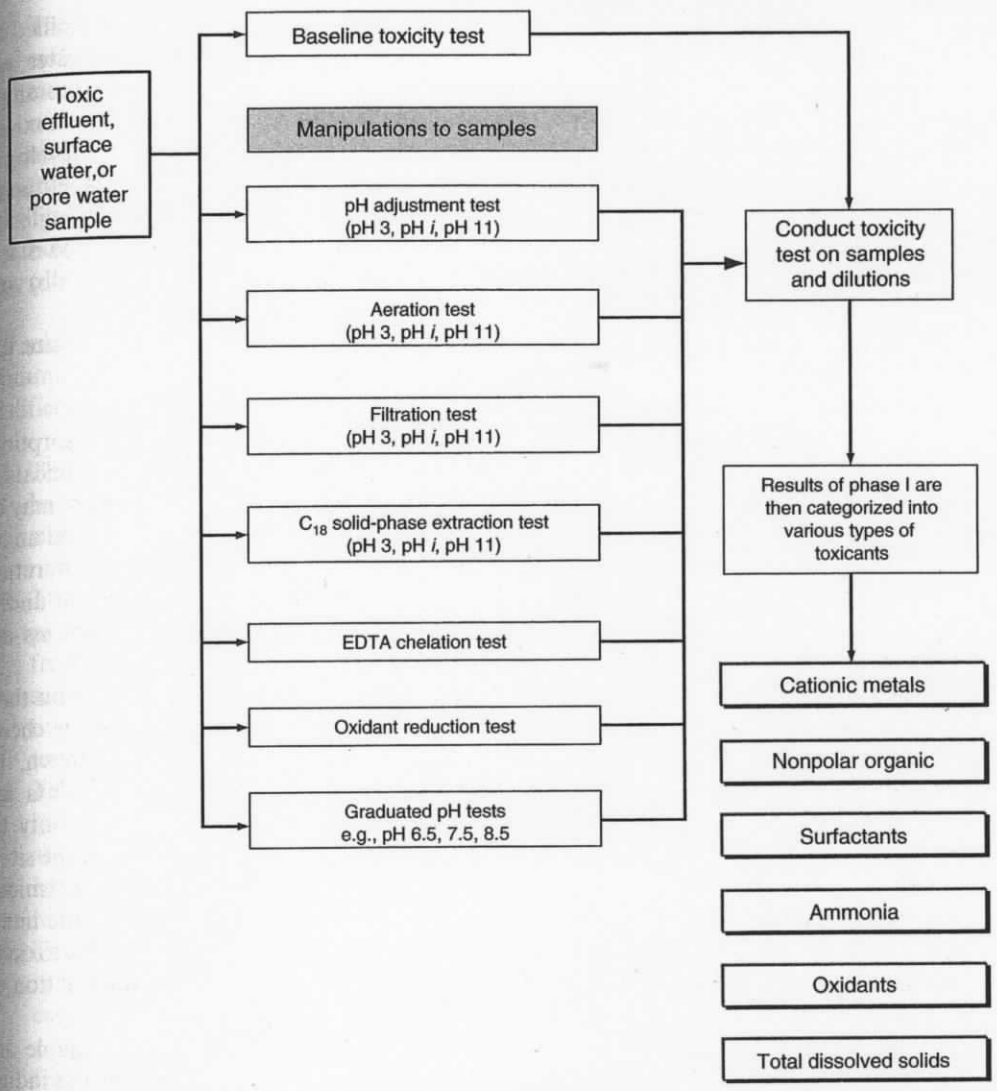

FIGURE 24.1 A logic diagram for toxicity identification evaluation (TIE) for acutely toxic aqueous samples. (From EPA (US Environmental Protection Agency), *Methods for Aquatic Toxicity Identification Evaluations: Phase I Toxicity Characterization Procedures*, 2nd ed., EPA-600/6-91-003, US Environmental Protection Agency, Duluth, MN, 1991a. With permission.)

1993a,b; Norberg-King et al. 2005). In TIE, the toxic constituents of a mixture are identified by removing components of a mixture and testing the residue, fractionating the mixture and testing the fractions, adding components of the mixture to background medium and testing those components, or other techniques (Figure 24.1).

24.5 MEDIA TESTS

The toxicity or other adverse properties of ambient media can be tested in at least three ways. In the least-used technique, contaminated biota are brought into the laboratory and tested. This technique is appropriate if the contaminant is persistent and bioaccumulated, or if it is

known to cause persistent injury. For example, herring eggs from areas exposed to spilled oil and from unexposed areas were brought into the laboratory, and their hatching rates and frequencies of abnormalities recorded (Pearson et al. 1995). By far the most common approach is to bring contaminated and reference media into the laboratory for toxicity testing. This is a very active area of ecotoxicology, and test methods have been developed for ambient waters, sediments, soils, and biota. Methods for testing aqueous effluents (Section 24.4) have been adapted to testing ambient waters and media (Norberg-King et al. 2005). Methods specifically recommended for use at contaminated sites in the United States and Canada may be found in Office of Emergency and Remedial Response (1994b) and Keddy et al. (1995).

In assessments of contaminated sites, testing the contaminated media from the site has several advantages relative to testing individual chemicals in laboratory media:

1. The bioavailability of the contaminants is realistically represented. Because of sorption, formation of complexes, and other processes that reduce the availability of a chemical for uptake by organisms, the toxic effects of a particular concentration of a chemical may be highly variable. In particular, standard single chemical toxicity tests are conducted under conditions that tend to maximize bioavailability, so toxicity values from the literature may be conservative. Media toxicity tests can reduce or eliminate this source of uncertainty by conserving the bioavailability of the contaminants to which organisms are exposed on the site.
2. The forms of the contaminants are realistic. The toxicity of chemicals depends on their form including the ionization states and co-ions for metals and other ionizable chemicals. Typically, the forms of contaminants at a site are unknown. Even when known, the predominant forms found at the site may not be those for which toxicity data are available. Media toxicity tests can reduce or eliminate this source of uncertainty by conserving the forms of the contaminants to which organisms are exposed on the site.
3. Combined toxic effects are elicited. Few sites are contaminated by only one chemical, and the toxic interactions of chemicals are seldom well known. In addition, the interactions depend on the form of the chemicals which is itself problematical. Media toxicity tests can reduce or eliminate this source of uncertainty by retaining the combination of contaminants in the forms and proportions that occur at the site.
4. The effects of contaminants for which few or no relevant test data are available are included. Ecotoxicological testing has focused on pesticides and metals, not the industrial chemicals found at many contaminated sites. Even for metals and pesticides, the taxa and responses of interest may not have been tested. Media toxicity tests greatly reduce or eliminate this source of uncertainty by including all contaminants to which organisms are exposed on the site in a test that has been chosen to represent the endpoint response.
5. The type of effects may be determined. The specific effects of the mixture may not be predictable from available knowledge of the effects of the components. The test can be designed to determine the occurrence of effects that are relevant to the assessment endpoint.
6. The spatial distribution of toxicity can be determined (Figure 20.3). The extent of the area to be assessed or remediated and the priority to be assigned to different sources or receptor ecosystems can be more appropriately determined on the basis of the distribution of toxicity than from the distribution of individual chemical concentrations.
7. Remedial goals may be determined. Toxicity can provide a better basis for defining media and areas to be remediated than chemical concentrations can.

8. The potential for achieving the level of anthropogenic effects specified in the assessment endpoint can be determined. In some cases, because of upstream or background contamination, it may be uncertain whether site remediation will significantly improve the ecological condition of the receiving system. Demonstrated toxicity of upstream water and sediments can provide a better basis for this determination than chemical concentrations can.
9. The efficacy of remedial actions can be determined. In many cases, toxicity provides a better basis for defining whether additional remediation is needed than chemical concentrations can.

For reasons such as these, media toxicity testing has been recommended by the EPA for use at contaminated sites (Office of Emergency and Remedial Response 1994c). However, the qualifiers in the statements above point to the following limitations in media toxicity testing:

1. The medium may be modified by collection and preparation for toxicity testing. This has been a particular concern for testing of sediments, which may lose their physical structure and oxidation state during collection, sieving, and storage. Soils and water may be modified as well.
2. The forms and concentrations of chemicals may be modified by sample collection and processing. These changes may result from the changes in the medium just discussed or from direct effects such as changes in speciation due to adjustment of pH or addition of salts to make the medium acceptable to the test species, or from loss of volatile chemicals to air or loss of chemicals from solution due to sorption to the walls of the sampling and testing containers.
3. The samples may be unrepresentative. This problem also occurs in sampling for chemical analysis, but may be more severe for media testing, because typically fewer samples are tested than are analyzed.
4. Most media toxicity tests have short durations and few response parameters are recorded relative to conventional chronic toxicity tests.
5. The cause of the toxicity is unknown. Toxicity may be due to one or more contaminants in the tested medium, so it may not be clear what remedial actions are needed. In some cases, apparent toxicity may be due to extraneous factors such as chemical or physical properties of the medium or disease. For example, it was necessary to UV sterilize water from the Oak Ridge Reservation for fathead minnow larval tests because of an unidentified pathogen.
6. Apparent toxicity may be due to the choice of inappropriate reference locations. For example, relatively rapid growth may be due to high nutrient levels in reference media rather than toxicity in site media.

These limitations do not negate the considerable advantages of media toxicity testing. The first three can be minimized by care in the collection and handling of samples and in the conduct of the tests. The fourth point requires analysis and interpretation of the results during the risk characterization, as with other test data.

The fifth problem requires applying TIE to contaminated ambient media as described above for effluents (Section 24.4). An example is the use of TIE to demonstrate that low concentrations of nickel were responsible for the toxicity to *Ceriodaphnia* of water from Bear Creek, Tennessee (Kszos et al. 1992). Methods are also being developed for sediments and pore water (Ankley and Schubauer-Berigan 1995; Burgess et al. 2000; Norberg-King et al. 2005). Standard TIE methods are not available for soils, but could be developed ad hoc. This may be most readily done by spiking background soil. For example, toxicity of soils from the

Lehigh Gap, Pennsylvania, to isopods were correlated with concentrations of several metals, but tests of soils spiked with individual metals showed that zinc was responsible (Beyer and Storm 1995). Extension of the TIE process to include other properties of tested media could solve the sixth problem.

Both control and reference media should be tested along with the contaminated media. Control media are laboratory media that are known to be appropriate for the test species; they support nearly maximal rates of survival, growth, and reproduction of the test species. Their characteristics are usually prescribed in standard test protocols. Reference media come from the vicinity of the contaminated site, and are physically and chemically similar to test media except that they do not contain the site contaminants. Reference media include waters and sediments collected upstream of the site or soils from the same soil series as the contaminated soils, but not contaminated by the site. If upstream reference media are contaminated by an upstream source or if local soils are contaminated by a source other than the site (e.g., historic use of an arsenical pesticide or atmospheric deposition from a smelter), it may be desirable to obtain reference samples outside the range of those sources (a clean reference as opposed to the local reference). The control tests determine whether the test was conducted properly using healthy organisms. The local reference tests provide the basis for determining how much toxicity the site adds to proximate media. If a separate clean reference is used, it provides the basis for determining whether the differences from controls are due to contaminants or to properties of the media such as pH or texture. For example, water from Poplar Creek on the Oak Ridge Reservation was toxic to Japanese medaka embryos, water immediately upstream was a little less toxic, and water several kilometers upstream, above a municipal waste water treatment plant, was not toxic (equivalent to controls).

As in any form of toxicology, the best evidence for toxic effects is provided by demonstration of an exposure–response relationship. This can be done by testing samples collected on a contamination gradient or by testing a dilution series. An obvious example of the former is sampling and testing waters in a gradient downstream of an effluent or contaminated site. Often, particularly for soils, contamination gradients do not occur on a site. In such cases, an exposure series can be created by diluting the contaminated medium with a clean reference medium. It is obviously important to ensure that factors such as nutrient levels or texture do not confound the toxic effects by carefully matching the dilution medium to the test medium. Finally, an exposure–response series can be established by spiking site or reference media with the chemicals of concern. In such studies, it is important to match the forms of the test chemicals to those of the site contaminants. For soils and sediments, it is also appropriate to age the media to establish more realistic bioavailability (Heiger-Bernays et al. 1997). In addition to establishing that toxicity is responsible for observed effects, an exposure–response relationship can be used to define remedial goals by establishing what level of the site's contaminant mixture has acceptably low levels of effects.

Conventionally, media test data are analyzed using hypothesis testing statistics. Responses of each tested medium are determined to be either statistically significantly different from those in reference or control media, or not. If an exposure–response relationship can be established by gradients or dilution series, a function may be fitted to the data and used to estimate exposure levels that cause prescribed levels of effects (LC_{50}, EC_{10}, etc.). If, as is nearly always the case, there is a mixture of contaminants, exposure must be expressed as concentration of a representative chemical or some metric of aggregate concentration such as total petroleum hydrocarbons. If there are no exposure gradients, assessors should at least report the level of effects relative to reference or controls in the tests of the individual samples and the associated variance among replicates (e.g., proportional mortality $= 0.22 \pm 0.12$) (Box 24.1).

BOX 24.1
Replication of Tested Media

Care must be taken when analyzing media toxicity data to understand the nature of the replicates. Often, a sample is taken from a water body or site, subsampled in the laboratory, and the subsamples used as replicates for testing. Such laboratory replicates incorporate variance among containers within a test but not variance in the material. If the intent is to characterize the toxicity of soil in an area of the site or in a stream, such laboratory replicates are pseudoreplicates. A separate sample should be taken for each replicate from the area or water body to be characterized. Note that the same consideration applies to replication in effluent testing.

In some cases, the site medium may be unsuitable for the test organisms to survive, grow, or reproduce. In those cases, one has the option of adjusting the medium or changing the test species. Adjusting the pH, hardness, or other physicochemical property is problematical because of the potential effects on the form, bioavailability, or toxicity of the contaminants. Such adjustments should not be performed when the medium properties are unsuitable due to properties of the waste. For example, leachate from the S-3 ponds on the Oak Ridge Reservation was highly acidic and had high metal concentrations, so it would have been inappropriate to adjust the pH of test waters from the receiving stream. However, in cases where the medium is naturally unsuitable for the test species, and the adjustment will not affect the chemical state of the contaminants, adjustments may be appropriate. Alternately, use of a test species that is appropriate to the medium is conceptually more appealing. Particularly if the chosen species is resident on the site, use of a species characteristic of the type of ecosystem that is being assessed increases the apparent relevance of the test results. However, a standard test species may not be available that is adapted to the site medium and it may be difficult to develop testing procedures for a nonstandard species.

Tests for specific media are discussed below and test protocols are summarized in accompanying tables. In general, the tests performed at contaminated sites follow standard protocols from the US EPA, ASTM, or other organizations. Standard media tests have the advantage that they are reasonably reliable, can be performed at reasonable cost by many laboratories, are acceptable to most regulators, and have known sensitivities. The most common deviation from standard protocols is the substitution of local species. However, some thought should be applied to determining whether nonstandard tests may be more appropriate. For example, neither the acute lethality tests nor the "subchronic" 7 d larval toxicity tests can detect effects on fecundity or early development of fish. Any chemical that has a primary mode of action that involves disruption of endocrine control of the formation of gametes or development of embryos would not be adequately tested by those methods. A reproductive test might be more appropriate (Ankley et al. 2001). Another example would be chemicals with very slow uptake kinetics that would not be adequately tested unless that test was long enough for equilibrium between the test organisms and the medium to be attained. In such cases, longer than standard test durations may be needed. Finally, the relationship of the test species and responses to the assessment endpoints must be considered. For example, if some property of a reptile or amphibian population is the assessment endpoint, none of the standard tests are suitable.

24.5.1 Contaminated Water Tests

The tests that were developed for testing of aqueous effluents are also commonly used for tests of ambient waters (Section 24.5 and Table 24.3). Specifically in the United States, they are

recommended for assessments of Superfund sites (Office of Emergency and Remedial Response 1994a) and surface waters (EPA 1991b). Assessors using tests of contaminated water must consider the variance in aqueous contamination. The three water samples used in a 7 d static-replacement test may vary considerably in their composition and toxicity, so effects may be due to one particularly toxic sample out of three or effects may not occur because of one particularly clean sample. Of course, further variance occurs at longer time scales. Hence, it is important to conduct analyses of water in conjunction with the tests. Further, it is important to consider the processes that contaminate the water and their influence on toxicity. If contaminant levels are likely to be elevated during low flows due to poor dilution, storm events due to runoff, spray drift following pesticide application, or episodic releases of effluents, water for testing should be collected at those times rather than on a schedule.

24.5.2 Contaminated Sediment Tests

Tests of contaminated sediments are in a less advanced stage of development than aqueous media tests. Relatively few protocols have been standardized (Table 24.4) and they have not been thoroughly validated against biosurvey data. However, short-term tests with marine and estuarine amphipods and polychaetes have been extensively used to evaluate the relative toxicity of coastal sediments (Long et al. 1995; MacDonald et al. 1996). Standard protocols for freshwater sediments in the United States are available for the amphipod *Hyalella azteca* and the midge *Chironomus tentans*. The selection of test organisms depends, in part, on their sensitivity to the site contaminants and tolerance to ecological conditions such as salinity and grain size. For example, *H. azteca* can be used in tests of estuarine ($\leq$15% salinity) but not marine sediments, whereas *Rhepoxynius abronius* can be used only in marine sediments.

TABLE 24.4
Standard Procedures Used to Test the Toxicity of Ambient Sediments. The Tests Listed Include those that have been Recommended by the EPA for Contaminated Sites

Species	Life Stage	Response	Duration (d)	Medium[a]
Chironomus tentans	Larvae	Mortality and growth	10	FW
Hyalella azteca	N/A	Mortality and growth	10	FW/ME[b]
Amphipod, marine sp.[c]	N/A	Mortality, emergence, and reburial	10	ME
Polychaetes[d]	Recently emerged juveniles and young adults	Mortality	10	ME
Polychaetes[d]	Recently emerged juveniles	Mortality and growth	20–28	ME

Source: Office of Emergency and Remedial Response, *Catalog of Standard Toxicity Tests for Ecological Risk Assessment*, EPA 540-F-94-013, US Environmental Protection Agency, Washington, DC, 1994a.With permission.

[a]FW = freshwater sediment; ME = marine or estuarine sediment; FW/ME = protocols are available for species from both media.

[b]*H. azteca* can be tested in estuarine sediments up to 15% salinity.

[c]The standard marine or estuarine amphipods in the United States are *Rhepoxynius abronius*, *Eohaustorius estuarius*, *Ampelisca abdita*, *Grandidierella japonica*, and *Leptocheirus plumulosus*.

[d]The standard marine or estuarine polychaetous annelids in the United States are *Neanthes arenaceodentata* and *N. virens*.

d Remedial Response
minated water must
s used in a 7 d static-
so effects may be due
se of one particularly
nce, it is important to
ortant to consider the
ontaminant levels are
due to runoff, spray
r for testing should be

pment than aqueous
24.4) and they have
hort-term tests with
ely used to evaluate
Donald et al. 1996).
re available for the
ection of test organ-
ts and tolerance to
. *azteca* can be used
whereas *Rhepoxynius*

The Tests Listed
ated Sites

Duration (d)	Medium[a]
10	FW
10	FW/ME[b]
10	ME
10	ME
20–28	ME

for Ecological Risk
1994a. With permission.
are available for species

ius, Eohaustorius estuarius,

es arenaceodentata and

While sediments are more stable than waters, seasonal changes in sediment chemistry (e.g., redox potential) may modify the bioavailability and toxicity of sediment-associated contaminants. For example, toxicity and bioconcentration tests of clams (*Mya arenaria*) were performed with a mixture of metals in water designed to simulate the interstitial water of Narraganset Bay (Eisler 1995). The mixture was lethal at simulated summer temperatures but not winter temperatures. Hence, it is advisable to conduct seasonal tests and to consider environmental characteristics that may modify toxicity.

Sediment tests are generally conducted using whole sediment samples and benthic infauna or epibenthic fauna. Sediment interstitial water can be tested using standard aquatic toxicity test methods. The pore water is extracted from the bulk sediment sample and, typically, used in tests with invertebrates. This approach can help identify the mechanisms of exposure for benthic infauna (i.e., respiration of contaminated pore water, ingestion of contaminated sediments, or both). This knowledge can be used to plan further sampling and interpret the results of other analyses (e.g., benthic invertebrate surveys). The disadvantage is that the extraction and testing processes can alter the form and bioavailability of the contaminants. Sediment elutriates may be formed by mixing sediments with water and then filtering or centrifuging. Elutriates may be tested to estimate effects of dredging, dredge spoil disposal, or other activities that suspend sediments in the water column. Guidance for collecting, handling, storing, and testing sediments can be found in Marine Protection Branch (1991) and Office of Water (2001).

24.5.3 Contaminated Soil Tests

Most tests of contaminated soils use seedlings of vascular plants or earthworms. Some standard soil tests are presented in Table 24.5, and the tests for single chemicals can be adapted to contaminated soil testing (Section 24.2.3). A review can be found in Wentsel et al. (2003).

When tests are performed in soil from a contaminated site, there is less need for normalization of the chemical concentrations than when using literature values. However, care must be taken to match reference soils to contaminated soils in terms of chemistry, texture, and nutrient status. Particularly for growth and reproduction endpoints, tests may be highly sensitive to soil properties. Therefore, it is desirable to test soils from multiple reference locations to estimate the natural variation. Where variation is significant, it can be reduced by normalization of concentrations. For example, soil metal concentrations were pH-normalized to reduce variation among locations at a metal mining and milling site in Anaconda, Montana (Kapustka et al. 1995). If site and reference soils have low organic matter or inorganic nutrient levels, it may be necessary to amend or fertilize them in order to support the test organisms, to achieve reasonable growth in the reference soils, or to bring site and reference soils to the same levels.

As with single chemical toxicity tests (Section 24.2), the choice of test organisms and toxicity endpoints should balance practicality of standard tests with the need to be relevant to the assessment endpoints. If possible, invertebrate species should be representative of assessment endpoints in function, taxonomy, trophic level, life history strategy, and route of exposure to toxicants (Spurgeon and Hopkin 1996). Tests of earthworms are relevant to the soil invertebrate community in most nonarid continental ecosystems. *Eisenia fetida*, the most common test organism in soil toxicology, has about average sensitivity to toxicants among earthworms (Laskowski et al. 1998a).

Most phytotoxicity tests of contaminated soil use crop species. For Superfund assessments, lettuce is listed as the standard species of the seed germination and root elongation assay, though the use of other species is permitted (Greene et al. 1988; Kapustka 1997). LeJeune et al. (1996) tested the tree crop, hybrid poplar, as a surrogate for spp. and *Populus* spp. at the

TABLE 24.5
Standard Procedures Used to Test the Toxicity of Ambient Soils

Type of Organism	Life Stage	Response	Duration (d)	Medium	Reference
Earthworm[1]	Adult	Mortality	14	Soil	Greene et al. 1988
Earthworm	Adult	Reproduction	35	Soil	ISO 1997
Plant[1]	Seed	Germination	5	Soil	Greene et al. 1988; Linder et al. 1992
Plant[1]	Seed	Root elongation	5	Elutriate	Greene et al. 1988; Linder et al. 1992
Plant	Seedling	Mortality and vegetative vigor	20–90	Soil	Linder et al. 1992
Plant	Seedling	Weight	45	Soil	Linder et al. 1992
Plant	Life cycle	Reproduction and growth	28–44	Elutriate	Linder et al. 1992

[1]Recommended by the EPA for use at contaminated sites, in Office of Emergency and Remedial Response, *Catalog of Standard Toxicity Tests for Ecological Risk Assessment*, EPA 540-F-94-013, US Environmental Protection Agency, Washington, DC, 1994a.

Clark Fork River, Montana. Many crops grow well in the laboratory, and they have replicable toxic effects, because seed of standard strains can be used.

Because ecological risk assessors are interested in the toxicity to native endpoint species, it may be advantageous to test them. For example, standard plant species have different root morphology, development patterns, and carbon allocation patterns, that may make results of a root elongation test more or less relevant to assessment endpoint species. Kapustka (1997) offers advice to the assessor who plans to use nonstandard test species. A few of his rules are: choose nominal performance standards (e.g., percent germination), characterize variability using reference soils, and use one or more standard species in addition to the nonstandard species. In general, organisms used in toxicity tests should be laboratory cultured or at least started from seeds or eggs to obtain uniformity of history, age, and physical condition. If the local species cannot be raised in the laboratory, they should be maintained in the laboratory for at least 2 weeks prior to testing, and control organisms should be maintained and observed beyond the end of the test to assure their good condition (Laskowski et al. 1998b).

Soil for toxicity tests should be sampled to an appropriate depth for the endpoint plant, invertebrate, or community. Appropriate depths would be the depths within which most exposure occurs. These are also depths within which most plant roots and invertebrate activity occurs. In one of the Clark Fork River studies described above, a deeper depth interval was used for tests with hybrid poplar than with alfalfa, lettuce, and wheat (LeJeune et al. 1996).

24.5.4 Ambient Media Tests with Wildlife

As with chemicals, contaminated media may be tested for oral, dermal, or inhalation exposures of wildlife. In general oral tests are most applicable for risk assessment purposes. Contaminated food, soil, or water from the site of interest may be collected and fed or orally administered to test animals. If dermal or inhalation exposure pathways are considered critical at a particular site, toxicity tests for these pathways may be performed. Standard methods for ambient media toxicity tests for wildlife do not exist. Methods for these tests may be developed ad hoc or by modification of standard chemical test methods (Section 24.2.4).

While ambient media toxicity tests with wildlife are not widely performed, examples do exist (Table 24.6).

Media toxicity tests for wildlife are rare, largely because of the expense of housing, feeding, and caring for adequate numbers of test animals. Obtaining sufficient contaminated food materials to maintain test animals for the duration of the study may also be difficult. For example, the mink toxicity study performed as part of the Clinch River Ecological Risk Assessment (Halbrook et al. 1999a) required more than 2000 kg of contaminated fish to maintain 50 mink for 6.5 months. In addition, many wildlife species reproduce only once a year and have multiyear generation times, so reproductive tests are time-consuming.

24.6 FIELD TESTS

The most direct approach to ecological toxicity testing is to treat real ecosystems. These tests may consist of treating and monitoring replicate areas of an ecosystem type (i.e., plot studies) or distinct ecosystems like the experimental lakes area in Canada (Schindler 1974, 1987; Schindler et al. 1985). If a site is already contaminated or disturbed, tests may involve caging, penning, or planting organisms along a gradient of contamination or disturbance, or at matched sets of contaminated or disturbed locations and reference locations. These approaches, termed field experiments, field-testing, or in situ testing, are relatively easy for immobile organisms such as plants and relatively distinct ecosystems such as ponds. They are more difficult for organisms that are mobile and forage for food and for ecosystems that are large or involve flowing water. Field-testing may be highly realistic, in that the organisms can be subjected to realistic conditions and variation in exposure. However, such studies are subject to the effects of variation among sites in conditions other than contamination and to loss of the study due to vandalism, predation, or extreme conditions. In addition, cage effects may modify the sensitivity of the organisms. Finally, field tests may involve deliberately injuring an ecosystem. Advocates of field-testing argue that microcosms or mesocosms provide replication at the expense of realism (Schindler 1998). While the US EPA no longer requires field tests for registration of pesticides, it does encourage field tests as part of postregistration monitoring (Tuart and Maciorowski 1997).

Some methods are intermediate between testing and monitoring. Dosing of replicate plots or ponds is clearly testing, and the sampling of naturally occurring organisms from contaminated or disturbed sites is clearly monitoring. However, the study of molluscs confined in contaminated and uncontaminated bays or of birds using nest boxes in areas with different levels of contamination includes some but not all aspects of a test. Organisms may be randomly assigned to locations, but the contamination or disturbance is not randomly assigned to those locations and the locations may not be replicated, so the study may be confounded. Such studies are included here for the pragmatic reason that they require manipulation of the system by an investigator.

24.6.1 Aquatic Field Tests

Aquatic field tests are most often performed by confining organisms in the water column or sediment of contaminated systems and reference systems (Chappie and Burton 2000). A good example of this technique is the use of caged mussels or clams to measure the uptake of contaminants and associated effects on survival and growth (Jenkins et al. 1995; Salazar and Salazar 1998; Donkin et al. 2003). These tests are sufficiently common and well developed to have a standard guide (ASTM E: 2122-01). Bivalves may be suspended in the water column, placed in trays on the sediment, or even placed in the intertidal zone. In situ tests of other

TABLE 24.6
Examples of Ambient Media Toxicity Tests to Evaluate Effects of Environmental Contaminants on Wildlife

Test Species	Reason for Test	Contaminants of Concern	Test Media	Toxicity Test Endpoint	Reference
Mink	Determine toxicity of Great Lakes fish to wild mink	Organochlorine pesticides, PCBs, dioxins	Diets containing carp from Saginaw Bay, MI	Mortality, reproduction, hematology, liver pathology, bioaccumulation	Heaton et al. 1995a,b
Mink	Determine toxicity of fish downstream of a US DOE facility to wild mink	PCBs, mercury	Diets containing fish from Poplar Creek, TN	Reproduction	Halbrook et al. 1999
Least shrew	Determine toxicity of metals in sewage sludge to secondary consumers	Cadmium, copper, lead, zinc	Diets containing earthworms from a sewage sludge-treated site	Growth, bioaccumulation	Brueske and Barrett 1991
Mallard, Ferret	Determine toxicity of weathered Exxon Valdez crude oil to seabirds and sea otters	Weathered crude oil	Weathered Exxon Valdez crude oil by capsules, gavage, or incorporated into diets	Mortality Food avoidance Organ pathology Reproduction	Stubblefield et al. 1995a–c
Mute swan, Canada goose, mallard	Determine toxicity of contaminated sediments to tundra swans	Lead	Coeur d'Alene River sediments in diet	Multiple physiological	Beyer et al. 2000

Source: Modified from Suter, G.W., II, Efroymson, R.A., Sample, B.E., and Jones, D.S., *Ecological Risk Assessment for Contaminated Sites*, Lewis Publishers, Boca Raton, FL, 2000. With permission.

aquatic organisms is limited by size, mobility, and food requirements. Small benthic invertebrates may be used to test sediment toxicity using trays with mesh bottoms and perforated covers (Chappie and Burton 1997; Tucker and Burton 1999). The suitability of fish for long-term exposures is limited by their mobility and food requirements. However, individuals of small species and young fish may be used in short-term exposures such as tests of storm runoff (Newbry and Lee 1984; Hall et al. 1988). Fish eggs in mesh bags may be a good alternative for longer exposures (Hiraoka and Okuda 1984). Similarly, amphibian eggs and larvae may be experimentally confined in the field (Linder et al. 2003). Such tests may reveal interactions among agents and conditions and address agents other than chemicals such as suspended sediment or temperature.

Field tests may also involve treating replicate aquatic ecosystems such as streams, small lakes, or ponds, or segments of an ecosystem. Such tests were commonly conducted in the past to determine effects of pesticides (Giles 1970; Jeffrey et al. 1986). Exposures of caged organisms may be conducted in conjunction with these ecosystem exposures (Clark et al. 1986). They can provide realistic exposure levels and conditions, observations of indirect effects and effects on properties of populations and ecosystems, and observations of recovery. However, these tests typically have few replicates and high variance among replicates.

24.6.2 Field Tests of Plants and Soil Organisms

Field-testing with soil organisms or communities is rare and methods are not well standardized. An example of field-testing is the placement of worms for 7 d in contaminated soil in plastic buckets buried at the locations where the soil was collected (Menzie et al. 1992). This study determined that highly toxic soils occurred as veins through site drainage areas. Carabid beetles have been tested in field pens on pesticide-contaminated soils (Heimbach et al. 1994). Field tests may be performed for processes as well as organisms. Effects on soil function may be tested in the field by using introduced substrates and monitoring their loss, by measuring respiration or nitrogen transformation, or by measuring enzyme activities. For example, the loss of tensile strength of buried cotton strips was used to test the effects of wood preservatives (Yeates et al. 1994). The most common method involves burying bags of leaf litter or crop residues in contaminated soils or, in natural systems, placing them on the litter layer. The soils may be at contaminated sites (Strojan 1978) or may be experimentally contaminated to test a pesticide or other chemical (Rombke et al. 2003). Field-testing methods for soils have been evaluated by Linder et al. (1992) and Wentsel et al. (2003).

Field tests of pesticides may involve spraying fields at expected application rates. Properties of plants, soil organisms, or soil functions may be measured and related to those rates. For example, the US EPA test guidelines for phytotoxicity (OPPTS 850.4025, 850.4300) include injuries and effects on production of natural vegetations, crops, or lawns. Soil or plant responses may also be determined incidentally in larger-scale field experiments for effects on wildlife.

24.6.3 Wildlife Field Tests

The primary advantage of wildlife field tests is that they allow ambient conditions to influence effects, and therefore may provide a more realistic measure of actual toxicity at a site or at sites with similar conditions. Due to the great mobility of most wildlife species, field tests are problematic for chronic exposures of most species and are generally suitable only for species with small home ranges. Some wildlife field tests involve creating penned areas of habitat that are treated to achieve defined contaminant levels or that have prior contamination. Since most such studies involve treating the pens, they are discussed above as mesocosms

Source: Modified from Suter, G.W., II, Efroymson, R.A., Sample, B.E., and Jones, D.S., *Ecological Risk Assessment for Contaminated Sites*, Lewis Publishers, Boca Raton, FL, 2000. With permission.

sediments in diet
contaminated sediments
to tundra swans
geese, mallard

(Section 24.3). However, the same approaches could be applied to a contaminated or experimentally treated site.

Tests of acute effects of pesticides on wildlife lend themselves to field tests; the treatment of a field results in realistic conditions and scale, and the movement of organisms on and off the field is realistic. Such tests may involve application of the substance to actual fields or forests followed by monitoring of birds, bees, or other nontarget organisms (OPPTS 1996a–d). Avian field tests of pesticides have served to reveal effects that were previously unknown, confirm effects that were suggested by observations, disprove effects that were suggested by laboratory studies, and demonstrate secondary effects such as reduced survival of young due to loss of insects (Blus and Henny 1997). Because these organisms are mobile and, unlike aquatic organisms, soil organisms, or plants, are not immersed in a contaminated medium, it is important to confirm exposure by analysis of gut contents, body burdens, or biomarkers (Balcomb et al. 1984). Dead organisms should be necropsied to determine the cause. Radio tagging allows investigators to determine the extent of use of the contaminated or disturbed area and to assure recovery of the test subjects.

Field tests of avian reproduction can be facilitated by using nest boxes to attract cavity-nesting birds. They may be attached to posts or trees within or around an area that is or will be contaminated or disturbed. Because cavities are frequently a limiting resource for cavity-nesting birds, nest boxes are likely to be occupied. Once the birds become established, effects on behavior and survival of adults and on the number, diet, growth, and development of young can be studied. Nest boxes are used by a variety of species including starlings, bluebirds, tree swallows, wood ducks, barn owls, and kestrels. Guidance for the use of nest boxes for the studies of starlings at contaminated sites is presented in EPA (1989). Nest boxes have been employed to evaluate risks to birds from application of insecticides to agricultural fields or forests (Robinson et al. 1988; Pascual 1994; Craft and Craft 1996), PCBs and heavy metals at a Superfund site (Arenal and Halbrook 1997), and lead along a highway (Grue et al. 1986).

24.7 TESTING ORGANISMS

Testing of biocontrol agents, generically engineered crops, and other exotic organisms is analogous to testing chemicals. Potentially susceptible organisms or ecosystems are exposed in the laboratory or field to the test organism at defined levels, and responses are recorded. In the United States, biocontrol agents must be tested for their ability to attack or infect nontarget organisms (OPPTS 1996g). Environment Canada has developed a set of ecological tests for new microbial agents (McLeay et al. 2004). Such tests are analogous to toxicity tests, including tests of plants, invertebrates, and vertebrates exposed orally, by injection, by inhalation, and by exposure to microbes in water. Reported responses include infection, pathogenicity, symptoms of toxins, and conventional responses (survival, growth, fecundity). In some cases, tests must be developed ad hoc. For example, testing for effects of *Bt* corn on monarch butterflies involved feeding studies of larvae on milkweed leaves with defined levels of pollen (grains/cm^2) (Stanley-Horn et al. 2001). Genetically engineered organisms such as *Bt* corn and enhanced *Rhizobium* are usually field-tested before approval (McClung and Sayre 1994). Such tests are essentially field trials of the organism in its agricultural use, but with monitoring of fate and potential effects on nontarget organisms. Because effects of organisms are more diverse than those of chemicals and because of their potential to multiply and spread, it is important to base tests of organisms on a careful problem formulation, which considers the possible activities of the organism with respect to species and ecosystems other than the targets.

24.8 TESTING OTHER NONCHEMICAL AGENTS

Because ecological risk assessment may be applied to any hazardous agent or process, various testing methods are needed to provide exposure–response relationships. Examples of hazards to be tested include harvesting methods, water storage and diversion, construction of utilities such as roads and pipelines, farming practices, and ecosystem management practices such as burning and mowing. Since tests are simply experimental applications of an agent, this implies the adaptation of ecological experiments to assure that the results define the relationship between the level of exposure and effects on endpoint entities and attributes. Examples include rain exclusion experiments to test effects of climate change (Yarie and Van Cleve 1996), watershed studies of forestry practices (Coweeta, Hubbard Brook, Walker Branch, etc.), forest exposures to CO_2 to determine effects of elevated atmospheric levels (Zak et al. 2003), and manipulation of dam-regulated hydrology (National Research Council 1999).

24.9 SUMMARY OF TESTING

Exposure–response tests are the core of ecological risk assessment. This is true because only controlled, replicated, and randomized exposures provide assurance that the observed associations between exposure and effect are causal. However, that assurance is limited to the test itself. The use of test results to estimate risks requires an extrapolation from test conditions to real, uncontrolled field situations, and this is discussed in the next three chapters.