26 Organism-Level Extrapolation Models

Life was certainly simpler in the old days... when we could evaluate risk with a safety factor.

Doull (1984)

In most cases, ecological risk assessments must be based on exposure–response data for species, life stages, levels of organization, and responses other than those specified by the assessment endpoints. In some cases, no exposure–response data are available for the agent of concern for any relevant organisms. For example, the endpoint is brook trout production and we have only a fathead minnow LC_{50} or only the structure of the chemical. Hence, it is necessary to use extrapolation models based on assumptions or statistical analyses to extrapolate to the endpoint species or community or to estimate parameters for population or ecosystem models (Chapter 27 and Chapter 28).

26.1 STRUCTURE–ACTIVITY RELATIONSHIPS

It is axiomatic that the biological effects of a chemical are a function of its structure. Hence, empirical models have been used to predict the pharmacological effects of drugs, the intended toxic effects of pesticides, and unintended toxic effects of these and other chemicals from their structural properties. These models, termed structure–activity relationships (SARs), are used to estimate effects when test data are unavailable. SARs may be qualitative, but the most useful models are quantitative (QSARs). (QSARs are also used to estimate fate-related properties; Chapter 22.)

SARs may be used to classify chemicals or to predict their toxicity. Chemicals may be classified in terms of whether they possess some property such as carcinogenicity, estrogenicity, or teratogenicity. This information may be used to reject a chemical during development or registration or to design tests by identifying potential test endpoints. Chemicals may also be classified in terms of their mode or mechanism of action (MoA; Chapter 7). If a chemical or class of chemicals is fitted by a QSAR that is associated with a particular MoA, it is likely to share that MoA. This information may be used as a basis for deciding that a concentration addition or dose addition model may be used to estimate the effects of mixtures of those chemicals (Chapter 8). MoAs include narcosis, respiratory uncoupling, and acetylcholinesterase inhibition. Predictions of toxicity from QSARs may be used directly to quantitatively estimate risks from untested or unreliably tested chemicals. Manufacturers use them to determine early in the process of chemical development whether a new chemical is likely to have significant toxic properties. They are used in a variety of regulatory contexts internationally, but are typically limited to screening applications (Cronin et al. 2003). For

example, in the US EPA, QSARs are used to screen industrial chemicals to determine whether testing is required (Nabholz et al. 1997; Zeeman 1995).

26.1.1 Chemical Domains for SARs

One of the more conceptually difficult aspects of SAR development and use is the identification of the domain, the range of chemicals from which a SAR should be derived and to which it may be applied. The most common approach is to identify a chemical class that is considered to be homologous such as aliphatic hydrocarbons or phenols. Congeneric chemicals are homologous sets that have a common functional group, termed a toxiphore. Examples of toxiphores include amine, hydroxy, sulfhydryl, and carboxyl groups. Examples of domains defined by the US EPA for screening industrial chemicals include aliphatic amines, dinitro benzenes, and phthalate esters (Nabholz et al. 1997). Artificial intelligence–based algorithms have been used, in place of expert judgment, to generate chemical classes based on toxiphores (Klopman et al. 2000).

Alternatively, domains may be defined in terms of the MoA of the constituent chemicals (Drummond et al. 1986) (Chapter 7). This is more reliable than use of chemical classes, because the members of a class may have different mechanisms of action (Russom et al. 1997). Hence, a chemical may belong to a well-defined class, but its toxicity may not be predicted by the model for that class. For example, phenols are often described as having a narcosis MoA (i.e., low toxicity), but most phenols have other modes of action that are more toxic to fathead minnows (Figure 7.1). The difficulty in this approach is that a study must be performed to identify the MoA. Some studies are capable of identifying multiple modes of action by reporting multiple physical and physiological responses in whole organisms (Bradbury et al. 1989). Others identify specific modes of action such as Ah receptor or estrogen agonists, often in vitro (Wenzel et al. 1997; Schmieder et al. 2000). In vitro approaches must be used with caution because of the potential for metabolism to alter the MoA. In addition, a chemical may have more than one MoA, and the identified MoA may be secondary to an unidentified mode.

26.1.2 Approaches for SARs

The basic approach is expert judgment. Judgment is always required, but in some cases it has been used to develop systems of rules for assessing the activity of chemicals (Walker 1993; Karabunarliev et al. 2002). Judgment may be used quantitatively, but more often is used qualitatively to predict that a chemical will have a particular property or to assign it to a category (e.g., likely to bioaccumulate). A quantitative example is the use of judgment to decide what tested chemical is most similar to the untested chemical; that chemical's activity data are then used as surrogates for missing data.

The most common approach to QSAR development is regression modeling. Toxic endpoint values such as fathead minnow 96 h LC_{50}s for a set of chemicals are regressed against some property of the chemical. The most common property in ecological QSARs is the octanol/water partitioning coefficient, K_{ow}. It is broadly useful because, for a set of organic chemicals with the same MoA, toxicity may be largely determined by the rate of uptake, which is in turn determined by hydrophobicity. A classic example is Konemann's (1981b) model of 14 d LC_{50} for neutral organic chemicals and fish:

$$\log(1/LC_{50}) = 0.87 \log K_{ow} - 4.87 \tag{26.1}$$

Subsequently, Veith et al. (1983) showed that for 96 h LC_{50}s, the model must be nonlinear at high values of K_{ow}, because the slow uptake of the high–molecular weight chemicals inhibits the attainment of a lethal internal level in an acute exposure:

$$\log LC_{50} = -0.94 \log K_{ow} + 0.94 \log (0.000068\ K_{ow} + 1) - 1.25 \quad (26.2)$$

These models describe toxicity by the baseline narcosis MoA in fish (Section 7.1).

A generic regression model for toxicity of organic chemicals is:

$$\log (1/C) = a(\text{hydrophobic}) + b(\text{electronic}) + c(\text{steric}) \quad (26.3)$$

where C is concentration and a, b, and c are fitted parameters (Hansch and Fujita 1964; Walker and Schultz 2003). Hydrophobicity is usually represented by K_{ow}; electronic properties may include charge, pKa, quantum chemical descriptors, or others; and steric properties include size and shape descriptors. Concentrations are usually expressed in terms of moles rather than mass, because, for a particular MoA, effects are more related to the number of molecules potentially reaching receptors than their mass.

26.1.3 State of SARs

The current state of practice in ecotoxicological SARs is to use simple statistical approaches to relate effects to external exposure metrics (usually concentration) as in toxicity tests (Walker and Schultz 2003). This practice has been summarized in the US EPA ECOSAR software, which contains more than 100 SARs for more than 40 chemical classes (http://www.epa.gov/oppt/newchems/21ecosar.html). Further development is needed to include more chronic effects and effects on more taxa, particularly terrestrial organisms. In addition, greater acceptance of QSARs for regulatory purposes will require more extensive and consistent quality assurance including greater transparency, better defined endpoints, molecular descriptors, and domains, and more mechanistic bases (Eriksson et al. 2003). One effort in that direction takes a multivariate approach employing a large number of molecular descriptors to classify chemicals as having similar or dissimilar modes of action and then develops QSARs for prediction of effective levels (Vighi et al. 2002).

Future developments are likely to involve the development of a computational ecotoxicology analogous to current practices in pharmacology. SARs will be used to derive the parameters of toxicokinetic and toxicodynamic models, which simulate the uptake, metabolism, distribution, excretion, and effects of chemicals (Yang et al. 1998). In advanced versions, molecular modeling of potentially toxic chemicals and the various receptors in organisms should allow the estimation of binding energies, which could be used to predict specific effects, as in drug design (Raffa 2001). This is more difficult for toxicologists than pharmacologists, because the receptors are seldom specified a priori. The Ah receptor, which is a target of dioxin-like chemicals, is one of the few exceptions (Mekenyan et al. 1996).

26.2 EFFECTS EXTRAPOLATION APPROACHES

Most analyses of effects begin with a small set of data related to a few responses of a few species and life stages and perhaps a few ecosystem responses. Somehow, assessors must extrapolate from those few data to the entities and responses that constitute the assessment endpoints. Assumptions, factors, and statistical models are most commonly used for this purpose. Increasingly, mechanistic models are used to extrapolate to population and ecosystem level endpoints (Chapter 27 and Chapter 28). In such cases, the effects parameters of the models must be estimated from the available data, and the same assumptions, factors, and statistical models are used for that purpose. Numerous and diverse extrapolation models have been developed, but their application has been somewhat haphazard, and there is no consensus about which are appropriate.

This section presents the major approaches for developing extrapolation models and then discusses the models that are used for particular media and taxa. Although each of the approaches is applicable to any assessment, different extrapolation approaches are used in different contexts because of the constraints of available data and differences in the traditions of the different groups of toxicologists.

26.2.1 Classification and Selection

It may be assumed that the endpoint species, life stages, and responses are equal to those in the most sensitive reported test or in the test that is most similar in terms of taxonomy or other factors. This process of classification and selection of test endpoints is the simplest and most commonly used extrapolation method. Sufficient similarity must be judged on the basis of some classification system. For example, plants are often classified by growth form, and the EPA has classified freshwater fish as warm-water and cold-water species (Stephan et al. 1985). However, species are most commonly classified taxonomically. Studies based on correlations of LC_{50}s of species at different taxonomic distances indicate that for both freshwater and marine fishes and arthropods, species within genera and genera within families tended to be relatively similar, which suggests that they can be treated as equivalent, given testing variance (Suter et al. 1983; LeBlanc 1984; Sloof et al. 1986; Suter and Rosen 1988). The same conclusion was reached for terrestrial vascular plants (Fletcher et al. 1990). Taxonomic patterns of sensitivity have been important in practice. For example, the observed levels of DDT/E in peregrine falcons and bald eagles did not appear to be sufficient to account for reproductive effects, until testing was done on a member of the same order (Lincer 1975). Other considerations in selection include the quality of the test, similarity of the test conditions to assessed field conditions, and relevance of measured responses.

In effect, this approach implies that the most sensitive or most relevant test organisms and conditions are surrogates for the endpoints in the field. Surrogacy is a concept with applicability to ecological risk assessments beyond toxicology. For example, when assessing a proposed introduction of a biocontrol agent, one would consider whether the tested potential nontarget hosts were adequate surrogates for the species, life stages, and exposure conditions that could occur in the field.

The advantage of this approach is its simplicity. One must choose the most appropriate test results in any case, and, by not applying any extrapolation model to the data, one avoids both effort and controversy. The disadvantage is that the available test data are seldom credible surrogates for the assessment endpoint response. Data selection is most defensible when based on a strong body of evidence. For example, one can use the threshold dose for developmental failure of chicken embryos to estimate risks to birds from dioxin-like chemicals, because a relatively large body of evidence indicates that it represents the critical response in birds, and chickens are a sensitive species (Giesy and Kannan 1998).

26.2.2 Factors

The next most common extrapolation method is to multiply or divide a test endpoint by a numerical factor. These factors are referred to as assessment factors, extrapolation factors, safety factors, and other terms. This extrapolation method may be treated as a formal extrapolation model

$$E_e = aE_t \tag{26.4}$$

where E_e and E_t are the effective exposure levels for the endpoint and test, respectively, and a can be statistically estimated (Section 26.2.5). However, in practice, the factors used in the

TABLE 26.1
Assessment Factors Used to Estimate Concern Levels in the Assessment of Industrial Chemicals by the US EPA Office of Pollution Prevention and Toxics. The Lowest Toxic Value is Divided by the Appropriate Factor to Set a Level of Concern for Exposures in the Environment[a]

Available Data	Assessment Factor
Limited (e.g., only one acute LC_{50} from a QSAR)	1000
Base set acute toxicity (fish and daphnid LC_{50} and algal EC_{50})	100
Chronic toxicity	10
Field test	1

Source: From Zeeman M.G., in G. Rand, ed., *Fundamentals of Aquatic Toxicology: Effects, Environmental Fate, and Risk Assessment*, Taylor & Francis, Washington, DC, 1995.

[a]A more complex set of factors, derived from this set, is used in Europe (CEC 1996).

regulation of chemicals are based on experience and judgment (Table 26.1) (OECD 1992, Zeeman 1995). The mathematical form is chosen for its simplicity, and multiples of ten are used to express order-of-magnitude precision. These factors have been often criticized, but they have been useful to regulatory assessors and have withstood legal scrutiny.

Sometimes, multiple factors are used. One might, for example, use a factor for interspecies differences, an acute/chronic factor, a laboratory/field factor, etc. These are nearly always treated multiplicatively

$$E_e = a_1 a_2 \cdots a_n E_t \tag{26.5}$$

for the n extrapolations incorporated. Multiplicative chains of safety factors imply that everything will go wrong together: the test species is maximally resistant relative to the endpoint species, there is a particularly large acute/chronic ratio, field conditions are particularly conducive to toxicity, etc. Because of this conservatism, such chains of factors are less often used than formerly. However, they have the advantage over single integrative factors of clarifying what extrapolations are incorporated, and their potential conservatism may be appropriate for screening assessments or when precaution is particularly desirable.

The primary advantage of factors is their ease. We can all divide by 10, 100, or 1000 in our heads. However, unlike simple data selection, factors allow adjustment for data inadequacies. The use of factors is widely accepted, and some have argued that there is no evidence that more sophisticated extrapolation models perform any better (Forbes and Forbes 1993; Forbes et al. 2001). Their primary disadvantages are that they are largely subjectively derived, and their use results in a value that is considered a safe level but is not clearly associated with a particular effect (Fairbrother and Kapustka 1996; Chapman et al. 1998). Hence, they are best used in screening assessments (Chapter 31).

26.2.3 Species Sensitivity Distributions

Species sensitivity distributions (SSDs), which were developed to estimate water quality standards that would protect some proportion of species, are increasingly used as extrapolation models in ecological risk assessments (Posthuma et al. 2001). SSDs are exposure–response models that are fitted to responses of species rather than organisms as in conventional toxicology (Chapter 23) (Figure 26.1). A percentile of the distribution of test endpoint values for

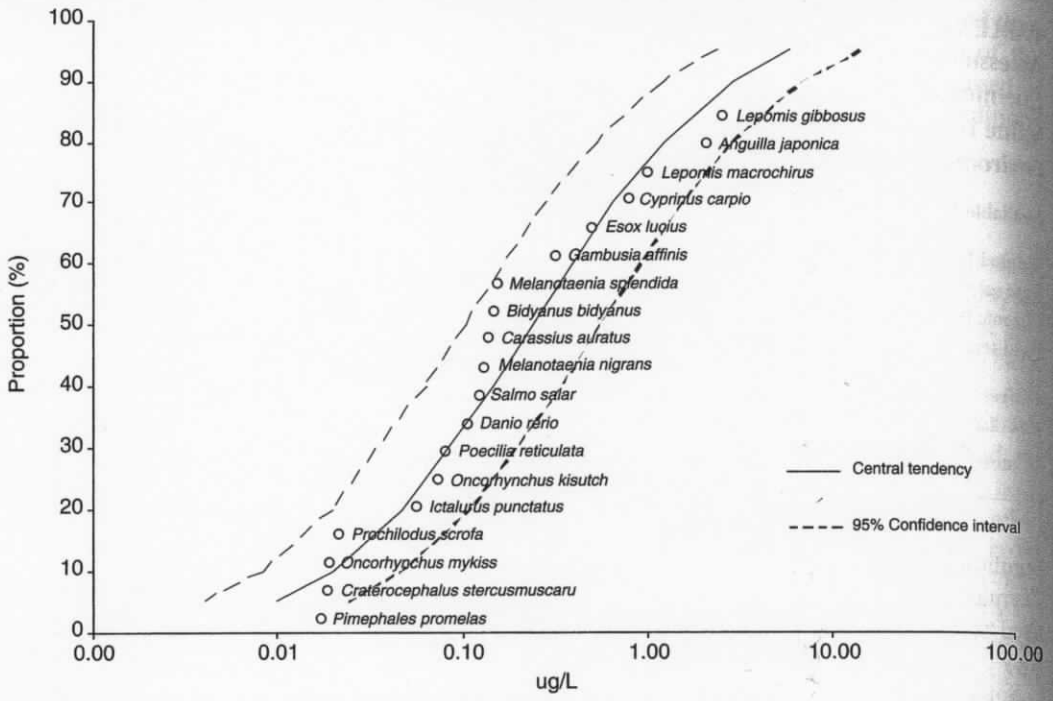


FIGURE 26.1 An example of a species sensitivity distribution (SSD), a logistic function relating the proportion of aquatic vertebrate species responding to copper in soft and warm water (<60 mg/L and >15°C). (Provided by Patricia Shaw-Allen. With permission.)

various species can be used to represent a concentration or dose that would affect that percentage of the exposed community. For example, if the 96 h LC_{50} values for fish exposed to a chemical are normally distributed (m_t, s_t), half of fish species in the field would be expected to experience mass mortality after exposure to concentration m_t within 96 h. This approach was developed for deriving water quality criteria independently in the United States (Stephan et al. 1985) and the Netherlands (Kooijman 1987). It has been repeatedly recommended as an ecological risk assessment technique (OECD 1992; Suter 1993a; Baker et al. 1994; Parkhurst et al. 1996a; EPA 1998a; Ecological Committee on FIFRA Risk Assessment Methods 1999a,b).

When SSDs are used to estimate levels of effects given an exposure level or to estimate the level of exposure corresponding to a level of effect, logistic or other functions are usually fitted to them. The choice of function makes relatively little difference if the data are well ordered (OECD 1992). Distributions may be fitted and percentiles calculated by any statistical software used for exposure–response modeling (Section 23.1.2). If used to support risk estimates based primarily on site-specific data or to support a causal analysis, an empirical distribution is simpler and adequate in most cases. Empirical distributions may even provide better numerical estimates, because many data sets are not fitted well by parametric functions (Newman et al. 2000).

The aggregation or partitioning of species in an SSD has been a topic of some debate. It is common practice in Europe to include all species, but the US EPA uses only multicellular animals to derive SSDs for water quality criteria, and others have advocated disaggregating as much as possible (e.g., fish, arthropods, other invertebrates, and algae). Aggregation of taxa in a common distribution provides more data with which to define the model. However, different taxa have different sensitivities, particularly for chemicals such as pesticides with

defined modes of action. In extreme cases, this leads to manifestly polymodal distributions (Figure 26.2). Hence, knowledge of MoA and taxonomic relationships should be used to determine whether and how to disaggregate a data set.

Because of their growing popularity, and because they are more technically sophisticated than data selection or factors, SSDs have been subjected to detailed critiques of their conceptual bases and practical implementation (Forbes and Forbes 1993; Smith and Cairns 1993). These concerns range from the practical (e.g., the minimum number of species) to the conceptual (e.g., reasonableness of using a set of single species tests to represent a biotic community), and are discussed at length in Suter et al. (2002).

An advantage of SSDs is that they make use of all relevant and reasonably standard test results. Further, an SSD can be readily interpreted as representing the distribution of responses of species in a community or a taxon. The chief limitations of this method are

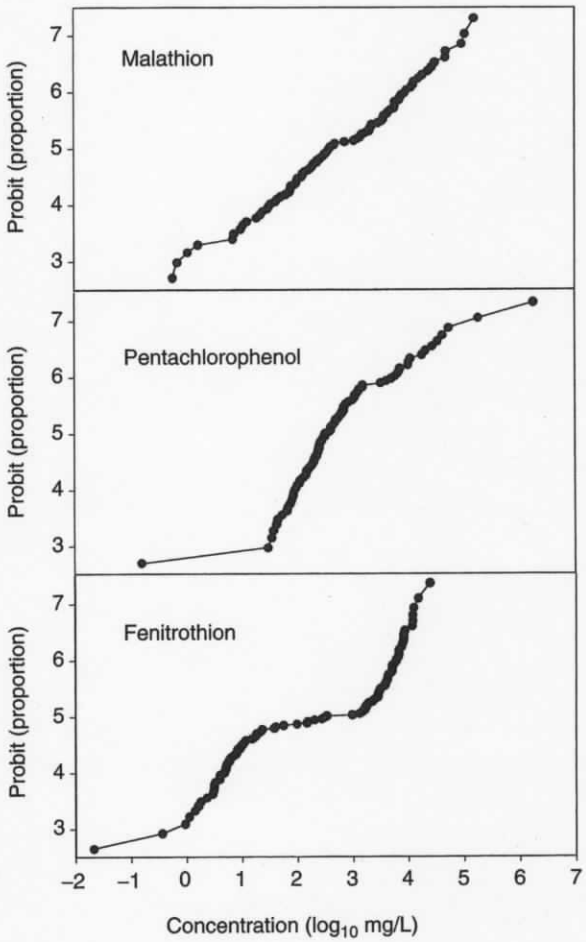


FIGURE 26.2 Species sensitivity distributions (SSDs) illustrating issues related to fitting standard functions. The Malathion SSD is linear on a log probit plot so it is fitted by the standard log-normal distribution. Pentachlorophenol has more highly sensitive and insensitive species than expected. Fenitrothion is bimodal, suggesting different modes of action or qualitatively different kinetics in different taxa. (From Newman, M.C., Ownby, D.R., Mezin, L.C.A., et al. in *Species Sensitivity Distributions in Ecotoxicology*, L. Posthuma, G.W. Suter II, and T.P. Traas, eds., Lewis Publishers, Boca Raton, FL, 2002. With permission.)

the requirement that enough species have been tested to define the SSD and that they be representative of the receiving community. When deriving water quality criteria, the EPA requires at least eight species from eight different families and they must be distributed across taxa in a prescribed manner, but regulatory assessments in the Netherlands may use as few as four species in an SSD. Relatively few chemicals have enough chronic toxicity data to establish the chronic SSD. Ingenious methods have been developed for developing SSDs when few data are available (Aldenberg and Luttik 2002; deZwart 2002). However, these approaches add significant uncertainty, so alternative extrapolation methods should be considered.

Another potential problem is that, if the media or the test conditions are variable and influential, the distributions will include extraneous variance. That is, the distribution is broader than the species sensitivities, because it includes variance due to conditions and test protocols. In fact, one extreme interpretation of SSDs is that they represent testing error, and the inherent differences in sensitivity among species are negligible (Van Straalen 2002a). Hence, attention must be paid to sources of variance and interpretation of SSDs. For aqueous toxicity extraneous variance can be low. The test methods and endpoints for aquatic toxic effects are reasonably consistent, so methodological variance should be relatively low. In addition, variance in test water chemistry is relatively low, particularly when hardness and pH normalization or speciation modeling are used for metals and ionizable compounds, so physical variance should be relatively low. For some chemicals, data are sufficiently abundant that SSDs can be derived for defined ranges of conditions (Shaw-Allen and Suter 2005). However, for both sediments and soils test data are sparse and the testing and survey methods and the endpoints are highly variable, the media have highly variable textures and chemistries, and reliable normalization methods are not available. Therefore, the physical and methodological variances may be significant contributors to the effects distributions in sediments and soils. The methodological variance is extraneous; the physical variance is an actual property of soils and sediments, and could be thought of as extraneous as well. However, if one takes an ecosystem perspective, the distributions resulting from the combination of biological and physical variance can be thought of as distributions of benthic ecosystem sensitivity, soil–plant system sensitivity, etc. It would be highly desirable to disaggregate those sources of variance by standardizing methods and by normalizing soil and sediment concentrations (Chapter 22).

Although SSDs were developed to address chemicals, they can be adapted to assess other agents. For example, the percentage of wildlife species behaviorally responding to aircraft overflights was related to slant distance to the aircraft (Efroymson and Suter 2001b).

SSDs may be interpreted in two ways (Suter 1993a, 1998a; Van Straalen 2002a). First, they may be interpreted as distributions of the probability that a species will be affected at a particular concentration. Hence, at a concentration of 100 μg/L, the probability of effects on any exposed aquatic species is 0.28 (Figure 26.3). Second, they may be treated as an estimate of the distribution of sensitivities of species in the exposed community. Hence, at a concentration of 100 μg/L the proportion of the community affected by the exposure is 0.28 (Figure 26.3). The results are a probability in the first case when the endpoint is effects on a population, but in the second case, when the endpoint is a community property, the result is deterministic.

The distinction may be clarified by analogy to exposure–response curves from conventional single species toxicity tests. The percentiles of those curves can be interpreted as probabilities of effects on individuals or as proportions of exposed populations. The former interpretation, which is used in human health risk assessments, is probabilistic, like the population-level interpretation of SSDs. The latter interpretation, which is more characteristic of ecological risk assessments but is also used in human health assessments, is deterministic, like the community-level interpretation of SSDs.

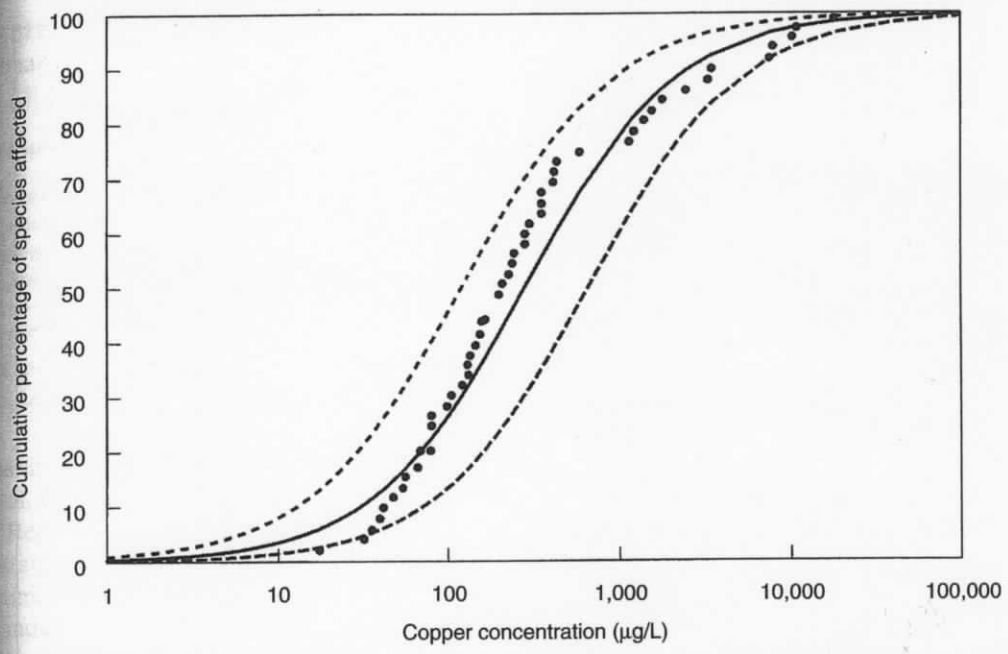


FIGURE 26.3 Cumulative species sensitivity distribution function for acute toxicity of copper. The curves are a logistic model fitted to the data points and upper and lower 95% prediction limits, generated using the Water Environment Research Foundation (WERF) software. (From Parkhurst, B.R., Warren-Hicks, W., Cardwell, R.D., Volosin, J., Etchison, T., Butcher, J.B., and Covington, S.M., *Aquatic Ecological Risk Assessment: A Multi-Tiered Approach*, Project 91-AER-1, Water Environment Research Foundation, 1996b With permission.)

In ecological risk assessments for the Oak Ridge Reservation, the aquatic assessment endpoints were defined at the community level and the endpoint properties included reductions in species richness or abundance (Suter et al. 1994). Hence, the appropriate interpretation is that the percentiles of the distribution are estimates of the proportion of species affected. In assessments of risks to aquatic communities, actual measurements of fish species richness and abundance as well as toxicity tests of ambient waters may be available. In such cases, the SSD is used to determine which chemicals are likely causes of any observed toxicity or community degradation rather than to estimate risks. For attributing cause, empirical distributions have been used and uncertainty analysis has not been judged important (Figure 26.4). However, if the risks are characterized on the basis of analyses of the SSD, uncertainty should be quantified.

The interpretation of SSDs is further complicated if uncertainty concerning the distributions is considered. Uncertainties concerning the percentiles of SSDs have been used to calculate conservative environmental criteria (Kooijman 1987; Aldenberg 1993; Aldenberg and Slob 1993) and to estimate risks as probabilities of effects (Parkhurst et al. 1996a,b). For a population endpoint, the percentiles of the distribution about the species sensitivity distribution are credibilities of a probability of effects on the endpoint population, i.e., we partition variability among species from uncertainty concerning that variable parameter. Consider Figure 26.3; the probability that a species will be affected at 100 µg/L is 0.27 with a credibility of 0.5 (the central line), but the upper 95% confidence bound indicates that there is a credibility of 0.025 (i.e., half of 5%) that the probability of effects is 0.46 or greater. This is conceptually equivalent to the output of a nested Monte Carlo simulation of an exposure

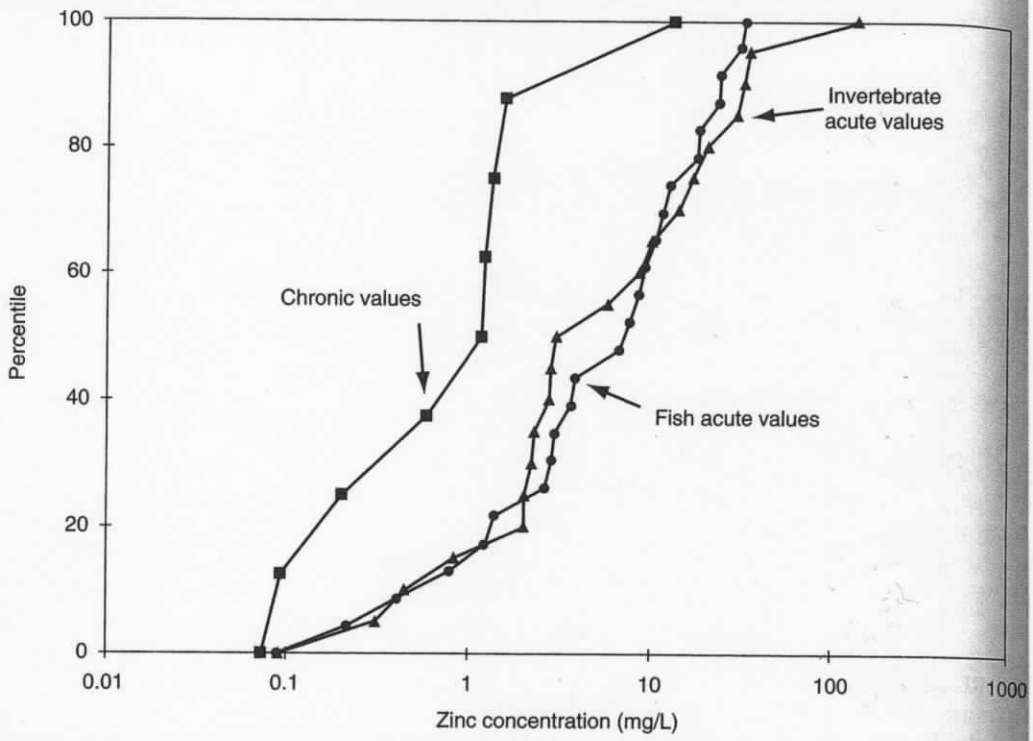


FIGURE 26.4 Empirical cumulative species sensitivity distributions for acute toxicity to fish, acute toxicity to aquatic invertebrates, and chronic toxicity to fish and invertebrates combined for zinc (From Suter, G.W., II, Efroymson, R.A., Sample, B.E., and Jones, D.S., *Ecological Risk Assessment for Contaminated Sites*, Lewis Publishers, Boca Raton, FL, 2000. With permission.)

model in which variability in exposure is partitioned from uncertainty. However, the uncertainty is simply the residual variance from the fitting of the distribution function to the toxicity data. This is a very incomplete estimate of the actual uncertainty (Suter 1993a). For a community endpoint, the percentiles of the distribution could be probabilities based on the variability among communities or credibilities based on the uncertainty concerning the distribution as a representation of community response. However, as in the population-level interpretation, the variance from the fit is an incomplete estimate of the actual uncertainty. The total uncertainty is a result of biases in the selection of test species, differences between laboratory and field sensitivities, and differences between the laboratory responses and the endpoint properties. Subjective uncertainty factors can be employed to estimate total uncertainty, but these should be based on a careful consideration of the data in the distribution and their relationship to the site-specific endpoint. A factor of 10 can be considered minimal. Research is needed to develop more objective estimates of the uncertainties concerning risk estimates derived from these distributions.

26.2.4 Regression Models

Regressions of one taxon on another, one life stage on another, one test duration on another, one level of organization on another, etc. can be used to extrapolate among taxa, life stages, durations, or levels of organization. This approach is extremely flexible and quantitatively rigorous but is seldom used. Regression models for aquatic extrapolations are presented in Table 26.2. More extensive discussions and examples of these methods can be found in the

TABLE 26.2
Linear Equations for Extrapolating from Standard Fish Test Species to All Bony Fish (Units are log μg/L)

Test Species	Slope	Intercept	*N*	Mean	F_1	F_2	PI[a]
Pimephales promelas	1.01	−0.30	354	2.77	0.45	0.0006	1.31
Lepomis macrochirus	0.96	0.17	500	2.52	0.49	0.0005	1.37
Oncorhynchus mykiss	0.99	0.29	480	2.42	0.38	0.0004	1.20
Cyprinodon variegatus	0.97	0.03	51	1.25	0.58	0.0085	1.49

Source: From Suter, G.W., II, *Ecological Risk Assessment*, Lewis Publishers, Boca Raton, FL, 1993a.

[a]PI, the 95% prediction interval at the mean, is log mean $Y \pm$ the number in this column.

literature (Suter et al. 1983, 1987; Barnthouse and Suter 1986; Sloof et al. 1986; Holcombe et al. 1988; Suter and Rosen 1988; Calabrese and Baldwin 1993; Mayer et al. 2004).

Regression models provide an alternative to SSDs when data are few. If a test endpoint for a standard test species is available, the distribution of the endpoint for all fish species can be estimated from the equations like those in Table 26.2 that regress all fish species against a standard test species for multiple chemicals (Barnthouse and Suter 1986; Suter et al. 1987; Holcombe et al. 1988; Suter and Rosen 1988). The equations estimate the mean of log LC_{50} for saltwater fish from *Cyprinodon variegatus* LC_{50} or for freshwater fish from the standard freshwater species. The 95% prediction interval (PI) at the mean is log mean $Y \pm$ PI. The PI is estimated from the variance in LC_{50}s for other species (Y) at a given LC_{50} for a standard test species (X_0):

$$\text{var}(Y|X_0) = F_1 + F_2(X_0 - X)^2 \tag{26.6}$$

Since the second term of the variance is relatively small, the PI at the mean is a reasonable estimate of the PI for all Y. That is, 95% of fish responses would be expected to fall within approximately ± 1.3 log units or approximately a factor of 20 of the log-normal mean fish response estimated from the equations.

A set of regression models for acute lethality to aquatic animals and terrestrial birds and mammals with supporting software, Interspecies Correlation Estimations (ICE), have been recently published (Asfaw et al. 2003). They are more convenient and are based on larger data sets than prior intertaxa regression models.

These intertaxa regression models were derived using data for a variety of chemicals. While this makes them generically useful, it also increases the variance. It is likely that the models could be made more precise by limiting them to a single mode of action, since relative sensitivities vary among chemical classes (Vaal et al. 1997).

26.2.5 Temporal Extrapolation of Exposure–Response Models

As exposure duration increases, the concentration or dose necessary to kill an organism declines to some minimum, the incipient lethal level. Hence, acute lethality tests can be used to estimate thresholds for effects in sustained exposures by extrapolating fitted exposure–response models to effectively infinite time (Mayer et al. 1994). A curve fitted to effects concentrations or doses at multiple times (e.g., 24, 48, 72, and 96 h LC_{50} values) will approach an asymptote or reach an effectively infinite duration such as the maximum life span of the organism. If one also extrapolates to a low response level (e.g., LC_{01}), the corresponding

concentration or dose may be interpreted as a safe level with respect to lethality for sustained exposures. Software—Acute to Chronic Estimation (ACE)—is available to perform such analyses using linear regression analysis, multifactor probit analysis, or accelerated life testing theory (Ellersieck et al. 2003). This approach has the advantage of estimating a threshold for chronic lethality without any chronic testing, but it requires that data for multiple acute durations be available. It also requires the assumption that the shape of the curve does not change at longer durations.

If data from multiple durations are not available, one may assume a relationship. The simplest is Haber's rule, which assumes that the product of dose or concentration (C) and time to effect (t) is a constant (k) for a particular effect. Therefore, for any exposure duration, the effective concentration or dose is:

$$C = k/t \tag{26.7}$$

This formula is commonly used for temporal extrapolation as well as to reduce the dimensionality of exposure–response models (Section 23.2.3), although it often poorly fits actual temporal data and should not be used for extrapolations beyond a narrow range. The need for this warning is obvious. Some concentrations are too low to cause an effect no matter how long the exposure. A proposed alternative is Ostwald's formula:

$$C = k/t^a \tag{26.8}$$

where a is a fitted constant. However, that formula requires either data fitting as in the ACE approach or a presumption that the exponent derived for a similar chemical with the same mode of action is applicable.

26.2.6 Factors Derived from Statistical Models

Most factors are derived by expert judgment based on experience or simple reviews of relationships among general types of data (Section 26.2.2), but factors may also be derived by data analysis and associated with a particular extrapolation. Sloof et al. (1986) used the PIs around regression models to derive uncertainty factors. Calabrese and Baldwin (1993) applied this approach to previously developed extrapolation models (Suter et al. 1983, 1987; Barnthouse and Suter 1986; Suter and Rosen 1988). Results for acute–chronic extrapolations for defined chronic responses and intertaxa extrapolations are shown in Table 26.3 and Table 26.4, respectively. The reader should note that this method retains only the highly conservative 90%, 95%, or 99% upper bound estimate of effects levels and not the best estimate.

The intertaxa extrapolations require some explanation. Suter et al. (1983) developed an approach for extrapolating between any test species and reference species that involved aggregation of species within taxonomic hierarchies. Using a large data set of aquatic acute toxicity data, congeneric species were regressed against each other and then aggregated; next, genera confamilial were regressed against each other and then aggregated; after that families within the same order were regressed against each other. This process continued up to a regression of the phylum vertebrata against the arthropoda. The increasing PIs on these regressions as the taxonomic distance increased was used to demonstrate that toxicological similarity is related to taxonomic similarity. Calabrese and Baldwin (1993) used a later version of the regressions for fish taxa to reduce the regressions and PIs to 95% and 99% uncertainty factors for each taxonomic relationship by calculating confidence intervals on the set of PIs for pairs of orders of fish (Table 26.5). Calabrese and Baldwin (1994) later suggested that these generic factors were applicable to taxa other than fish, including humans. For

TABLE 26.3
Uncertainty Factors for Extrapolations from Acute Lethality to Specific Chronic Effects in Fish

			Uncertainty Factors		
			Confidence Interval		
X Variable	*Y* Variable	*n*	90%	95%	99%
LC_{50}	Hatch EC_{25}	31	26	50	198
LC_{50}	Parent mort EC_{25}	28	18	32	106
LC_{50}	Larval mort EC_{25}	89	18	31	93
LC_{50}	Eggs EC_{25}	42	32	64	228
LC_{50}[a]	Fecundity EC_{25}	26	26	50	206
LC_{50}[a]	Weight[b] EC_{25}	37	28	53	188
LC_{50}[a]	Weight/egg EC_{25}	14	91	246	2247
Mean			34	75	467
Weighted mean			27	55	265

Source: From Calabrese, E.J. and Baldwin, L.A., *Performing Ecological Risk Assessments*, Lewis Press, Boca Raton, FL, 1993. With permission.

[a]Regression analysis from Suter et al. (1987).
[b]Decrease in weight of fish at end of larval stage.

TABLE 26.4
Taxonomic Extrapolation: Means and Weighted Means for the 95% and 99% Prediction Intervals (PIs) for Uncertainty Factors Calculated by Calabrese and Baldwin (1994)[a]

			Uncertainty Factor	
X Variable	*Y* Variable	*n*	95% PI	99% PI
	Taxonomic extrapolation: species within genera			
Salmo clarkii	*S. gairdneri*	18	9	13
S. clarkii	*S. salar*	6	6	10
S. clarkii	*S. trutta*	8	6	8
S. gairdneri	*S. salar*	10	7	11
S. gairdneri	*S. trutta*	15	4	5
S. salar	*S. trutta*	7	5	8
Ictalurus melas	*I. Punctatus*	12	5	7
Lepomis cyanellus	*L. macrochirus*	14	6	9
Fundulus heteroclitus	*F. majalis*	12	6	8
Mean			6	10
Weighted mean			6	7
	Taxonomic extrapolation: genera within families			
Oncorynchus	*Salmo*	56	5	6
Oncorynchus	*Salvelinus*	13	4	5
Salmo	*Salvelinus*	56	5	7
Carassius	*Cyprinus*	8	4	6
Carassius	*Pimephales*	19	7	9

Continued

TABLE 26.4 (Continued)
Taxonomic Extrapolation: Means and Weighted Means for the 95% and 99% Prediction Intervals (PIs) for Uncertainty Factors Calculated by Calabrese and Baldwin (1994)[a]

			Uncertainty Factor	
***X* Variable**	***Y* Variable**	***n***	**95% PI**	**99% PI**
Cyprinus	*Pimephales*	10	7	10
Lepomis	*Micropterus*	30	8	11
Lepomis	*Pomoxis*	8	9	13
Cyprinodon	*Fundulus*	12	6	8
Mean			6	8
Weighted Mean			6	8
	Taxonomic extrapolation: families within orders			
Centrarchidae	*Percidae*	47	10	14
Centrarchidae	*Cichlidae*	6	4	6
Percidae	*Cichlidae*	5	13	24
Percidae	*Esocidae*	11	9	13
Atherinidae	*Cyprinodontidae*	32	7	9
Mugilidae	*Labridae*	12	55	78
Cyprinodontidae	*Poecillidae*	12	3	5
Mean			14	21
Weighted mean			13	18
	Taxonomic extrapolation: orders within classes			
Salmoniformes	*Cypriniformes*	225	20	27
Salmoniformes	*Siluriformes*	203	39	51
Salmoniformes	*Perciformes*	443	12	16
Cypriniformes	*Siluriformes*	111	11	15
Cypriniformes	*Perciformes*	219	32	43
Siluriformes	*Perciformes*	190	63	83
Anguiliformes	*Tetraodontiformes*	12	13	18
Anguiliformes	*Perciformes*	34	25	34
Anguiliformes	*Gasterosteiformes*	8	16	24
Anguiliformes	*Atheriniformes*	46	9	12
Atheriniformes	*Cypriniformes*	7	501[b]	786[b]
Atheriniformes	*Tetraodontiformes*	46	13	17
Atheriniformes	*Perciformes*	148	25	33
Atheriniformes	*Gasterosteiformes*	36	20	27
Gasterosteiformes	*Tetraodontiformes*	8	20	30
Gasterosteiformes	*Perciformes*	33	32	43
Perciformes	*Tetraodontiformes*	34	25	34
Mean			24	32
Weighted mean			26	35

[a]Values in this table are similar to, but differ from, those in Barnthouse et al. (1990) due to differences in the algorithm used, particularly the use of ordinary least squares regression by Calabrese and Baldwin (1994).
[b]Not included in calculations.

ınd 99% Prediction
aldwin (1994)[a]

ainty Factor	
% PI	99% PI
7	10
8	11
9	13
6	8
6	8
6	8
10	14
4	6
13	24
9	13
7	9
55	78
3	5
14	21
13	18
20	27
39	51
12	16
11	15
32	43
63	83
13	18
25	34
16	24
9	12
)1[b]	786[b]
13	17
25	33
20	27
20	30
32	43
25	34
24	32
26	35

ifferences in the algorithm
994).

TABLE 26.5
Upper 95% Uncertainty Factors Calculated for the 95% and 99% Prediction Intervals in Table 26.4

	Prediction Interval	
Level of Taxonomic Extrapolation	**95%**	**99%**
Species within genera	10.0	16.3
Genera within families	11.7	16.9
Families within orders	99.5	145.0
Orders within classes	64.8	87.5

Source: From Calabrese, E.J. and Baldwin, L.A., *Environ. Health Perspect.*, 102, 14, 1994. With permission.

example, when extrapolating between a mouse test and equivalent effects on a mammalian carnivore (order Carnivora), one would divide the mouse test endpoint by 64.8 to be 95% certain of including the carnivore species 95% of the time (Table 26.5).

An alternative approach was developed for the calculation of tier II water quality values (Chapter 29). These values were derived by applying resampling statistics to the data sets used to derive water quality criteria to obtain distributions of the ratios of the lowest concentration in a small sample of toxicity data to the criteria values. Factors were derived from these ratios that should protect 95% of aquatic invertebrate and fish species with 80% confidence. This method is best used to develop conservative screening benchmarks.

26.2.7 Allometric Scaling

After factors, the type of quantitative extrapolation model used most commonly by human and wildlife toxicologists is allometric scaling. These models are based on the assumption that all members of a taxon have the same response to a chemical, but they differ in size and in processes that are related to size. The most commonly used allometric model is a power function of weight:

$$E_x = aW^b \tag{26.9}$$

where E_x is the effective dose or concentration at some organism weight W. This model has been adopted by toxicologists because the metabolism and excretion of drugs and other chemicals are approximated by that function. The EPA and others have used the three-fourth power for piscivorous wildlife (EPA 1993f; Sample et al. 1996c). Allometric scaling may be applied to aquatic species (Patin 1982), but it is used almost entirely for wildlife extrapolations. Reviews of the theory and application of allometric scaling are provided by Fairbrother and Kapustka (1996), Davidson et al. (1986), and Peters (1983).

Allometric scaling is simple to apply, and it has a stronger scientific basis than uncertainty factors. If a toxic dose (D) and the body weights of both the test and endpoint species are known and an appropriate scaling factor is selected, the toxicity value for the wildlife species may be calculated (Sample et al. 1996c):

$$D_w = D_t\left(\frac{bw_t}{bw_w}\right)^{1-b} \tag{26.10}$$

Confidence in allometric scaling is limited because current models are based on a few chemical classes (i.e., mammalian values are based primarily on drugs, and avian values are based primarily on cholinesterase-inhibiting insecticides). In addition, avian models are based on acute lethality. Because allometric scaling factors can vary widely among chemicals (Mineau et al. 1996), and because the toxic mode of action differs for acute and chronic exposures to the same chemical, the current practice of applying the same scaling factors for all chemicals and types of exposures may produce inaccurate estimates (Fairbrother and Kapustka 1996).

26.2.8 Toxicokinetic Modeling for Extrapolation

Toxicokinetic modeling, which is used to estimate body burdens and internal exposure levels (Section 22.9), also provides a means of extrapolating among species or life stages based on differences in physiology and the volumes of organs and other compartments. Toxicokinetic models are used to extrapolate rodent test data to humans (Clewell et al. 2002), but have seldom been used in ecological risk assessments. The conceptual approach is shown in Figure 26.5. It begins with the results of a conventional toxicity test, such as a laboratory rat reproductive test, expressed as the administered dose in mg/kg/d of the chemical of concern that causes a prescribed effect, such as 20% reduction in the number of viable pups. A toxicokinetic model is then used to estimate the corresponding internal exposure for the female rat. This may be the concentration in a particular compartment such as blood but is more likely to be a whole body concentration (mg/kg body weight). The internal concentration may be a peak concentration for a single dose or the equilibrium dose for a continuous exposure. This must be converted into an equivalent effective internal dose for females of the endpoint species (e.g., the internal concentration causing 20% reduction in the number of viable mink pups). Toxicodynamic models could be used to represent the induction of effects by the internal exposure (Section 23.3), but in practice, even in human health risk assessments, the effective internal concentrations are typically assumed to be equal. A toxicokinetic model for the endpoint species is then used to convert the internal concentration to an administered dose. This dose can then be related back to ambient concentrations in food or abiotic media using exposure models (Section 22.9).

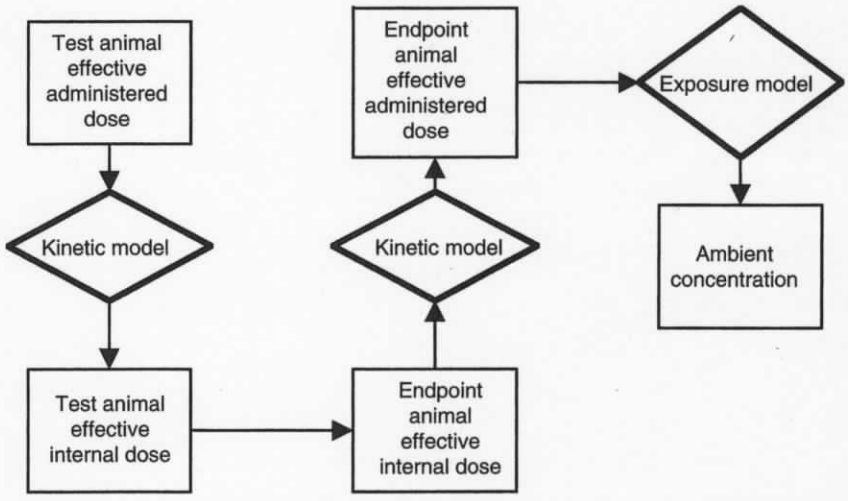


FIGURE 26.5 A process for using a toxicokinetic model to extrapolate between species.

The primary problem with this approach is that it is data-intensive, and most of the physiological parameters needed for most species are unavailable or poorly developed. However, despite data limitations, Fairbrother and Kapustka (1996) suggest that the use of even simple toxicokinetic models may significantly increase accuracy of interspecies extrapolations. One solution to data limitations is to use allometric relationships to estimate the needed parameters (Clewell et al. 2002). Another strategy is to focus on a few parameters that are likely to vary among species or organisms in a way that influences toxicity. For example, variance in the toxicity of chemicals to fish may be largely controlled by weight, the ratio of gill area to weight, and, for hydrophobic chemicals, by fraction of weight composed of fat (Lassiter and Hallam 1990). That is, when exposed to hydrophobic chemicals, big, fat, sluggish fish are most likely to survive.

Toxicodynamic models that relate internal exposure to response (Section 23.3) also have the potential to contribute to extrapolation among species and life stages. However, they are much less developed for that purpose than toxicokinetic models.

26.2.9 Multiple and Combined Approaches

These various potential methods for extrapolating toxic effects among species and life stages have traditionally been applied independently and not in a systematic manner. Fairbrother and Kapustka (1996) suggested that less reliance be placed on a single approach (e.g., allometric models) for all species and chemicals and that multiple approaches be applied to the problem of wildlife extrapolation.

To address all of the important extrapolations in an assessment, it is often necessary to combine approaches. For example, SSDs are used to extrapolate from individual species to communities, but the test endpoints used in the SSDs often do not represent the assessment endpoint. Hence, the individual species values for the model may be adjusted. For example, in calculating the acute ambient water quality criteria, the US EPA applies a factor of 2 to the 5th percentile of the SSD of acute values to extrapolate from 50% mortality to a small percentage (Stephan et al. 1985).

26.3 EXTRAPOLATIONS FOR PARTICULAR BIOTAS

26.3.1 Aquatic Biota

If, as is often the case, the endpoint property for the aquatic biota is species richness or diversity, SSD is an obvious choice of extrapolation model. Modeling results show that continuous exposure to concentrations equal to the chronic value (CV) for a species can cause extinction of that species (Barnthouse et al. 1990). Therefore, the proportion of species for which the CV is exceeded by long-term exposures can be assumed to approximate the proportion of species lost from the community. In addition, because toxicity data are relatively abundant for aquatic organisms, it is often feasible to derive such distributions for individual chemicals. As discussed above, the SSD approach is widely accepted, because it is used for the derivation of water quality criteria. If responses are known to be a function of water chemistry, the individual test endpoints should be normalized to appropriate water chemistry before defining the distribution.

If the endpoint is a property of a particular population rather than a community, the extrapolations using SSDs are performed or interpreted differently. SSDs are still useful because they can be interpreted as probability distributions for effects on an individual species (Section 26.2.3).

Alternatively, one can extrapolate between species by using the appropriate intertaxa regression models or the uncertainty factors derived from them (Table 26.4 and Table 26.5).

That is, if one wanted to predict the toxicity of a chemical to brook trout (a salmonid) from test data for fathead minnow (a cyprinid), one could divide by 20 to be 95% certain of not underestimating the sensitivity of brook trout (or any other salmonid). If the desired taxonomic regression is not available, the appropriate generic factor (which would be 26 in this case of an interorder extrapolation) would be applied. These two approaches for estimating effects on particular species or taxa (SSDs and taxonomic regressions) have different weaknesses, and it is not clear which works better in practice. However, the taxonomic regressions and the factors derived from them require test data for only one species, so they are more generally useful. The factors are conservative and may estimate effects levels that are below background. For estimation of probabilities of effects, one should use the original regression models to estimate means and variances (see Table 7.4 in Suter 1993a).

Allometric models may also be used to extrapolate to specific aquatic species (Patin 1982; Newman and Heagler 1991; Newman et al. 1994). This approach has not been accepted in practice, because taxonomic differences have been perceived to be more important than size. At least, it seems likely to be useful for extrapolation among taxonomically similar species.

Acute–chronic extrapolations may be made with regression models or factors. Acute–chronic regression models for aquatic fish and invertebrates are presented in Suter (1993a), and factors derived from them are presented in Table 26.3. These factors are based on including the CV or EC_{25} with 95% or 99% confidence. Alternatively, CVs can be estimated with 80% confidence of not overestimating their value using a factor of 18 (Host et al. 1991). Calabrese and Baldwin (1993) recommend generic 95% and 99% uncertainty factors of 50 and 200 for acute–chronic extrapolations, based on the weighted means in Table 26.3. Analysis of a European data set yielded various acute/chronic factors including a 90% confidence LC_{50} to no observed effect concentration (NOEC) factor for organic chemicals and aquatic animals of 24.5 (Lange et al. 1998). Any of these factors are adequate if one is trying to conservatively estimate a chronically toxic concentration of a chemical in a screening assessment or to support an assessment based primarily on other lines of evidence. If acute lethality data are available for at least three exposure durations, extrapolation to low response rates and long durations can be used to estimate lethal thresholds (Section 26.2.5). If compelled by circumstances to estimate risks to aquatic organisms using only an LC_{50}, one of the regression equations in Suter (1993a) or Sloof et al. (1986) with its associated uncertainty should provide a better estimate of the chronic effects threshold than a conservative generic factor.

In some cases, multiple extrapolations are required including those between taxa and life stages. Such multiple extrapolations may be dealt with by chains of factors or by chains of regression models (Barnthouse et al. 1990; Calabrese and Baldwin 1993; Suter 1993a). Thresholds for lethality obtained by temporal extrapolation (Section 26.2.5) may be converted to estimates of thresholds for reproductive and other sublethal effects by multiplying by a factor of 0.1 (Ellersieck et al. 2003). However, methods for multiple extrapolations are not explained in detail here.

Mechanistic models, particularly the biotic ligand model (Section 22.2), are beginning to be used in aquatic ecological risk assessments. They can be used to extrapolate among water chemistries and, potentially, among species and life stages.

26.3.2 Benthic Invertebrates

Species or community sensitivity distributions can be derived for toxicity of individual chemicals to benthic invertebrates based on published test results. In the case of exposure of benthic invertebrates to sediment pore water, the effects distributions are the same as the SSDs for aquatic biota. The use of aqueous data to evaluate effects on benthic species is based

on data suggesting that benthic species are not systematically more or less sensitive than water column species (EPA 1993a).

In the case of exposure of benthic invertebrates to chemicals in whole sediment, the effects distributions are for species/sediment combinations and community/sediment combinations. This is necessary because it is not possible to adequately control for the effect of sediment characteristics, including co-contaminants in field-collected sediments, on toxicity. The most prominent examples of effects distributions for benthic invertebrates are those used to derive screening benchmarks for sediment-associated biota (Long et al. 1995; MacDonald et al. 1996). The effects in those distributions include taxa richness, diversity, density, mortality, growth, respiration, behavior, and suborganismal effects. As a result, those distributions only indicate an unspecified level of an unspecified effect. This is adequate for screening purposes, but not for definitive risk characterization. For definitive assessments, such nonspecific distributions can be parsed into distributions of thresholds for specific effects. For example, Jones et al. (1999) developed distributions of community-level effects and lethality from the sediment toxicity data presented in MacDonald et al. (1996) and Long and Morgan (1991).

26.3.3 Wildlife

Wildlife risk assessors have followed health risk assessors in using uncertainty factors (e.g., Banton et al. 1996, Sample et al. 1996c, Hoff and Henningsen 1998). As in health risk assessments, uncertainty factors are used to account for specific extrapolations such as interspecies, acute–chronic, laboratory–field, lowest observed adverse effect level/no observed adverse effect level (LOAEL–NOAEL). A general extrapolation model for wildlife has been proposed by Hoff and Henningsen (1998):

$$D_w = D_t/(UF_a * UF_b * UF_c * UF_d) \tag{26.11}$$

where D_w represents the estimated critical chronic dose for an endpoint wildlife species, and D_t is the literature-derived toxicity value for the test species. UF_a accounts for intertaxon variability and can range from 1 if the test and wildlife species are the same to 5 if the test and wildlife species are in the same class but in different orders. Uncertainty in study duration is represented by UF_b, which ranges from 1 to 15 for the range from chronic to acute. UF_c accounts for the type of toxicity data available and ranges from 0.75 for NOELs to 15 for severe or lethal effects ($\gg ED_{50}$). Finally, UF_d addresses other modifying factors (e.g., species sensitivity, laboratory–field extrapolation, intraspecific variability) and may range from 0.5 to 2. Hoff and Henningsen (1998) recommend reporting quantitative risk results only if total $UF < 100$. For total $UF > 100$, only qualitative (e.g., presence/absence, low, medium, high) estimates of risk should be reported. As with other uses of multiplicative factors, this model suffers from inappropriate error propagation and poorly defended factors. However, it is equivalent to current practice in human health risk assessment.

Allometric scaling has been commonly applied to wildlife as well as humans. Factors of 0.66 to 0.75 have been used for extrapolating from laboratory test species to humans and wildlife (EPA 1992c, 1993e; Sample et al. 1996c). Use of either the 0.66 or 0.75 scaling factor is conservative for humans and most endpoint mammalian wildlife in that large species such as deer are estimated to be more sensitive than the rodents, which are typically used in mammalian toxicity testing. However, small wild mammals are estimated to be less sensitive than laboratory rats or dogs. Common avian test species such as chickens and mallard ducks are much larger than most birds.

Use of allometric models for birds with the same exponents as mammals was supported by allometric models of avian physiology and pharmacology (Peters 1983; Pokras et al. 1993). In

contrast, allometric regression analyses of 37 pesticides on 6 to 33 species of birds found that for 78% of chemicals the exponent was greater than 1 with a range of 0.63 to 1.55 and a mean of 1.1 (Mineau et al. 1996). Because of that mean and because scaling factors for the majority of the chemicals evaluated were not significantly different from 1, a scaling factor of 1 appears to be appropriate for interspecies extrapolation among birds (Sample and Arenal 1999). However, because of the apparently chemical-specific character of scaling factors, Sample and Arenal (1999) provided such factors for 138 chemicals for acute lethality to birds and 94 to mammals. In the absence of a scaling factor for the chemical of interest or for a similar chemical, they recommend generic factors of 1.2 for birds and 0.94 for mammals. Scaling factors should be developed for chronic exposures and for effects other than lethality.

SSDs have been recommended for assessment of risks to birds (Baril et al. 1994; Ecological Committee on FIFRA Risk Assessment Methods 1999b). However, sufficiently large data sets are limited primarily to acutely lethal doses of pesticides. When data are insufficient, safety factors derived from SSDs may be used (Baril et al. 1994). These factors are the geometric means, across chemicals, of the ratios of LD_{50}s for common test species to the TLD_5 (the 5th percentile of the avian SSD) (Table 26.6).

Regression models are also potentially useful for wildlife. Intertaxa regressions are available for avian and mammalian wildlife species and families (Asfaw et al. 2003; Mayer et al. 2004). Because there are no avian test results for most chemicals, regressions of available avian data against rat data allow the use of the much more abundant data for laboratory rats. For example, regressions of ring-necked pheasant LD_{50}s and mallard duck LD_{50}s against rat LD_{50}s for 24 organophosphate cholinesterase-inhibiting pesticides were used to assess risks to birds from disposal of organophosphate cholinesterase-inhibiting chemical warfare agents (Sigal and Suter 1989). The result for mallards was:

$$\log \text{mallard } LD_{50} = 1.33(\log \text{rat } LD_{50}) - 0.58 \tag{26.12}$$

($r^2 = 0.47$) indicating that mallards are more sensitive than rats to these chemicals. Pheasants, however, were about equal in sensitivity to rats.

26.3.4 Soil Invertebrates and Plants

Regression models were developed for pairs of plant taxa by Fletcher et al. (1990). For each of 16 chemicals, EC_{50}s of between 7 and 36 plant species were compared. The variation in sensitivity ranged from 3.5-fold for linuron to 316-fold for picloram. Out of almost 300 chemical–plant species combinations, 59% of EC_{50}s varied by less than a factor of 5 from other EC_{50}s for the same chemical. Plants that were closely related, taxonomically, had similar sensitivities to the same chemical. No trends in the relative sensitivity of various species, genera, or families to different chemicals were observed. In this study interspecies variability in toxicity was much higher than the variability associated with the extrapolation from greenhouse to field. Hence, when assessing risks to a particular plant species, it is apparently more important to use data for a similar species than for similar conditions. However, these results are based on foliar applications of herbicides.

Since assessment endpoints for soil invertebrates and plants usually include community properties, SSDs would seem appropriate. The OAK Ridge National Laboratory (ORNL) benchmarks for exposure to contaminants in soil used SSDs for earthworms and plants (Efroymson et al. 1997a,b) (Figure 26.6). Because the variance of toxicity among soils can be significant and cannot be factored out, soil type is a source of variability in these distributions. Hence, the points in Figure 26.6 are species–soil type combinations, and the distribution is the distribution of effective concentrations across species and soils. As a

TABLE 26.6
Safety Factors for Commonly Tested Bird Species. An LD_{50} for One of These Species Divided by the Geometric Mean Safety Factor Should Be Protective of 95% of Avian Species with the Indicated Reliability

Safety Factor	Pheasant	Mallard	Bobwhite	Japanese Quail	Red-Winged Blackbird	Starling	House Sparrow	Common Grackle	Rock Dove	Red Partridge	Grey Partridge
Geometric mean	16.8	10.9	15.2	17.1	5.87	19.8	10.7	9.26	13.1	21.6	10.3
Maximum	298	113	141	174	18.7	1250	43.9	48.42	55.2	87.8	79.8
Minimum	2.00	2.12	2.40	3.10	2.28	3.76	2.13	2.13	3.51	10.5	3.58
n	22	22	16	22	22	22	22	22	22	7	7
%Reliability[a]	41	36	38	50	41	41	45	50	50	43	29

Source: Modified from Baril, A., Jobin, B., Mineau, P., and Collins, B.T., *A Consideration of Inter-Species Variability in the Use of the Median Lethal Dose (LD_{50}) in Avian Risk Assessment*, No. 216. Hull, PQ, Canadian Wildlife Service, 1994. With permission.

[a]Percentage of chemicals for which the mean safety factor is sufficient to obtain 95% protection.

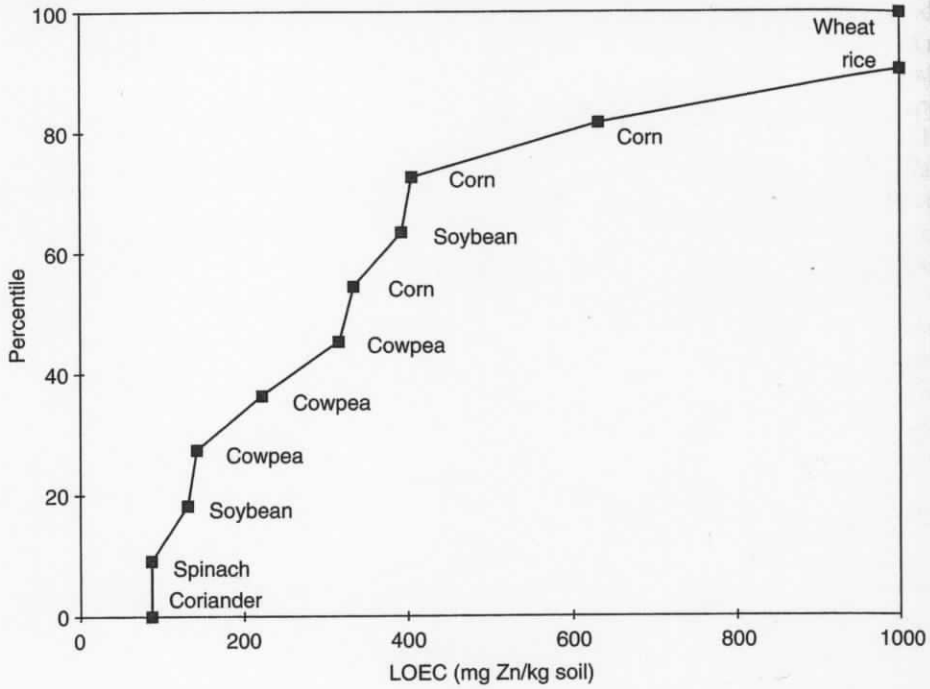


FIGURE 26.6 Cumulative distribution of LOECs for plants exposed to zinc in soil. Effects included are changes in the mass of whole plants of plant parts. (From Suter, G.W., II, Efroymson, R.A., Sample, B.E., and Jones, D.S., *Ecological Risk Assessment for Contaminated Sites*, Lewis Publishers, Boca Raton, FL, 2000. With permission.)

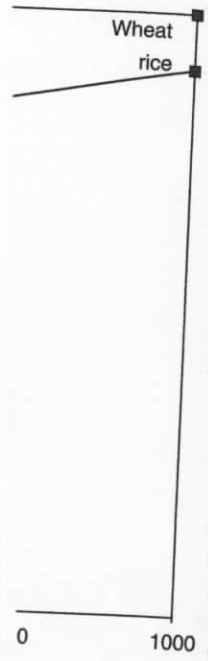
result, single plant species that are tested under different conditions have different LOECs in the distribution. An untested species in a particular soil may be assumed to be a random draw from the distribution, or the distribution may represent the proportion of species in a plant community that is likely to be affected by a particular concentration of a chemical given uncertainty concerning the influence of soil type. That uncertainty is reduced in the Netherlands by normalization to reference soil (Sijm et al. 2002).

26.3.5 Soil Processes

In the Netherlands, distributions of toxicity values for microbial processes and enzymatic activity are used, along with SSDs, to derive regulatory values for soil (Crommentuijn et al. 2000; Sijm et al. 2002). Studies of the same process in different soils are treated as separate observations, i.e., each soil ecosystem is considered equivalent to a species in an SSD. The NOEC and ECx concentration are normalized to standard soil to minimize the effects of soil chemistry on bioavailability and toxicity.

26.3.6 Water Chemistry

The properties of ambient and test waters such as salinity, pH, and hardness can influence the forms of chemicals to which organisms are exposed and the sensitivity of aquatic organisms to the exposure. The differences between test waters and waters to be assessed may be addressed in various ways. The simplest is to use data from tests conducted in waters that are similar to site water, but this raises the question of sufficient similarity. The question is particularly



oil. Effects included are
oymson, R.A., Sample,
Lewis Publishers, Boca

e different LOECs in
to be a random draw
of species in a plant
of a chemical given
luced in the Nether-

sses and enzymatic
Crommentuijn et al.
treated as separate
ies in an SSD. The
e the effects of soil

s can influence the
uatic organisms to
may be addressed
that are similar to
on is particularly

stark for the case of extrapolation from freshwater to saltwater. Few toxicity data are available for saltwater relative to freshwater, but there are few test data from freshwater sediments relative to estuarine sediments. Analyses have shown that the differences are small relative to the differences among species (Klapow and Lewis 1979; Hutchinson et al. 1998; deZwart 2002), but these differences may be important (Stephan et al. 1985; Hutchinson et al. 1998).

The second approach is to perform a standard test in site water and in standard test water and use the ratio of the results to adjust a water quality criterion or other test data for the effects of site water chemistry. Guidance for deriving and applying this water effect ratio is provided by the US EPA (Office of Science and Technology 1994, 2001).

A third approach is to model the influence of water chemistry. One example is the use of empirical models to adjust metal toxicity data for water hardness using regressions of LC_{50}s against test water hardness (Stephan et al. 1985; Pascoe et al. 1986). Metal speciation models have been used with the biotic ligand model to adjust the toxicity of certain metals for differences in the chemistry of freshwaters (Section 22.2).

26.3.7 Soil Properties

The risk assessor should be aware that bioavailability in soil from the contaminated site may be substantially different from the bioavailability in published soil tests. As stated in Section 22.4.1, aged organic chemicals are typically less available and less toxic to biota than organic chemicals freshly added to soil in published toxicity tests (Alexander et al. 1995); thus the toxicity at the contaminated site may be overestimated if a published toxicity test of a chemical freshly added to soil is emphasized too heavily in the assessment. The risk assessor can make adjustments to observed toxic concentrations to account for differences in soils or chemical speciation. The variance in toxicity among natural soils may be reduced by normalizing the test soil concentrations to match normalized site soil concentrations (Section 22.3) (Sijm et al. 2002). Or free metal activities in soil solution may be estimated, potentially improving the precision of toxic thresholds for plants, soil invertebrates, or microbial processes (Sauvé 2001). The assessor may be more liberal in including tests in screening assessments (e.g., in the derivation of screening benchmarks) than in definitive assessments. In definitive assessments, soil type and chemical speciation should be factors in decisions about the acceptability of data.

26.3.8 Laboratory to Field

Many studies have been conducted attempting to relate conventional laboratory toxicity test results to responses in the field. Unfortunately, most of them have been intended to test the validity of laboratory toxicity data rather than to generate extrapolation models. The simplest formulation of the validation problem is, do classifications of field sites as impaired or unimpaired correspond to the classifications predicted from standard laboratory tests? The results of these attempts at validation are ambiguous at best, largely because field studies do not provide "true" results as a standard for comparison. Field experiments and biological surveys are highly inconsistent in their design and endpoints; they are pseudoreplicated, poorly replicated, or unreplicated; they typically include only one season; although they often include multiple responses, they miss many taxa, attributes, and processes; and they tend to be insensitive, so that when effects are found they are often severe (Neuhold 1986; Chapman 1995; LaPoint 1995; Luoma 1995; deVlaming and Norberg-King 1999). Because mechanistic understanding is usually lacking, it is unclear whether deviations from expectations are due to failure of the laboratory tests as predictors or to factors that are not relevant

to evaluating the validity of a test. In addition, biological surveys are subject to confounding by differences among sampling locations other than the laboratory-tested toxicants.

Despite the difficulties, reviews of comparisons of laboratory and field studies typically conclude that when ambient dilution water or sediment is used, laboratory toxicity is usually indicative of field effects, and even tests with standard laboratory water are generally related to field effects (deVlaming and Norberg-King 1999; Long 2000). When laboratory and field results disagree, it is not clear that the laboratory test is erroneous. Even when they agree, it is not clear that the agreement is due to actual mechanistic correspondence. For example, field effects may be due to low dissolved oxygen rather than toxicity. Better comparisons are possible when consistent endpoints are used (e.g., percentiles of SSDs for invertebrates in laboratory and ditch mesocosm tests), but even then generalizations about the validity of laboratory tests are elusive (van den Brink et al. 2002).

An alternative approach to the relationship between laboratory and field test results is to simply regress the former against the latter. Sloof et al. (1986) regressed NOECs from aquatic mesocosm tests ($NOEC_e$) against the lowest reported single species acute LC_{50} or EC_{50} and against the lowest reported single species chronic no effect concentration ($NOEC_s$) for the same chemical and obtained:

$$\log NOEC_e = -0.55 + 0.81 \log LC_{50} \tag{26.13}$$

$$\log NOEC_e = 0.63 + 0.85 \log NOEC_s \tag{26.14}$$

The units are μg/L, the correlation coefficients are 0.77 and 0.85, respectively, and the arithmetic-scale PIs are ±0.86 and 0.35, respectively. The authors concluded that the most sensitive responses in the mesocosms are more sensitive than the most sensitive acute lethality test but less sensitive than the most sensitive single species chronic test. Emans et al. (1993) derived a model similar to Equation 26.14, using different data selection criteria. They concluded that organisms in field conditions responded at concentrations similar to those that affect species in the laboratory. Although these models would not accurately predict effects in particular cases, they could be used to suggest the approximate range within which effects would be expected in field systems. They also suggest that although more species are exposed in the field and indirect effects and complex exposures occur in the field that do not occur in simple laboratory aqueous tests, the chemicals that are most toxic in the laboratory are also most toxic in the field and thresholds for effects are not greatly different, so toxicity in the ecosystem context is not completely unpredictable.

When developing empirical models of field effects, it is important to select field data that are measures of the assessment endpoint, because different community and ecosystem responses have very different sensitivities. Even different community metrics derived from the same data set can give very different relative sensitivities. For example, the sediments of the Louisianian province were most affected of three EMAP coastal provinces based on a benthic index score, least affected based on species richness, and intermediate based on infaunal abundance (Long 2000).

One can certainly imagine more sophisticated and potentially more predictive approaches to empirically modeling the relationships between laboratory and field responses. In particular, the use of chemical speciation models or toxicokinetic exposure models to normalize field and laboratory contaminant concentrations, the use of tests with more life stages and responses, and the use of body burdens as measures of exposure are likely to improve predictions. However, more progress may be obtained using laboratory data with mechanistic models of exposure (Chapter 22), population dynamics (Chapter 27), and ecosystem processes (Chapter 28).

26.4 SUMMARY

Since the first edition of this book was published in 1992, the situation has changed little with respect to extrapolation models for estimating risks to attributes of nonhuman organisms. The use of SSDs has become common, but otherwise most assessments still rely on selection of most relevant data and the occasional factor. There is no consensus that this actually represents good practice or that a particular alternative is preferable. Looked at positively, it gives assessors considerable freedom to select the approach that seems best to them, or to develop new approaches.