APPENDIX C
SUPPLEMENTAL GUIDANCE ON LITERATURE SEARCH

APPENDIX C
SUPPLEMENTAL GUIDANCE ON LITERATURE SEARCH

A literature search is conducted to obtain information on contaminants of concern, their potential ecological effects, and species of concern. This appendix is separated into two sections; Section C-1 describes the information necessary for the literature review portion of an ecological risk assessment. Topics include information for exposure profiles, bioavailability or bioconcentration factors for various compounds, life-history information for the species of concern or the surrogate species, and an ecological effects profile. Section C-2 lists information sources and techniques for a literature search and review. Topics include a discussion of how to select key words on which to base a search and various sources of information (i.e., databases, scientific abstracts, literature reviews, journal articles, and government documents). Threatened and endangered species are discussed separately due to the unique databases and information sources available for these species.

Prior to conducting a literature search, it is important to determine what information is needed for the ecological risk assessment. The questions raised in Section D-1 must be thoroughly reviewed, the information necessary to complete the assessment must be determined, and the purpose of the assessment must be clearly defined. Once these activities are completed, the actual literature search can begin. These activities will assist in focusing and streamlining the search.

C-1 LITERATURE REVIEW FOR AN ECOLOGICAL RISK ASSESSMENT

Specific information. During problem formulation, the risk assessor must determine what information is needed for the risk assessment. For example, if the risk assessment will estimate the effects of lead contamination of soils on terrestrial vertebrates, then literature information on the effects of dissolved lead to fish would not be relevant. The type and form of the contaminant and the biological species of concern often can focus the literature search. For example, the toxicity of organometallic compounds is quite different from the comparable inorganic forms. Different isomers of organic compounds also can have different toxic effects.

Reports of toxicity tests should be reviewed critically to ensure that the study was scientifically sound. For example, a report should specify the exposure routes, measures of effect and exposure, and the full study design. Moreover, whether the investigator used accepted scientific techniques should be determined.

The exposure route used in the study should also be comparable to the exposure route in the risk assessment. Data reported for studies where exposure is by injection or gavage are not directly comparable to dietary exposure studies. Therefore, an uncertainty factor might need to be included in the risk assessment study design, or the toxicity report should not be used in the risk assessment.

To use some data reported in the literature, dose conversions are necessary to estimate toxicity levels for species other than those tested. Doses for many laboratory studies are reported in terms of mg contaminant/kg diet, sometimes on a wet-weight basis and sometimes on a dry-weight basis.

That expression should be converted to mg contaminant/kg wet bodyweight/day, so that estimates of an equivalent dose in another species can be scaled appropriately. Average ingestion rate and wet body weight for a species often are reported in the original toxicity study. If not, estimates of those data can be obtained from other literature sources to make the dose conversion:

Dose = (mg contaminant/kg diet) × ingestion rate (kg/day) × (1/wet body weight (kg)).

If the contaminant concentration is expressed as mg contaminant/kg dry diet, the ingestion rate should also be in terms of kg of dry diet ingested per day.

Exposure profile. Once contaminants of concern are selected for the ecological risk assessment, a general overview of the contaminants' physical and chemical properties is needed. The fate and transport of contaminants in the environment determines how biota are likely to be exposed. Many contaminants undergo degradation (e.g., hydrolysis, photolysis, microbial) after release into the environment. Degradation can affect toxicity, persistence, and fate and transport of compounds. Developing an exposure profile for a contaminant requires information regarding inherent properties of the contaminant that can affect fate and transport or bioavailability.

Bioavailability. Of particular importance in an ecological risk assessment is the bioavailability of site contaminants in the environment. Bioavailability influences exposure levels for the biota. Some factors that affect bioavailability of contaminants in soil and sediment include the proportion of the medium composed of organic matter, grain size of the medium, and its pH. The aerobic state of sediments is important because it often affects the chemical form of contaminants. Those physical properties of the media can change the chemical form of a contaminant to a form that is more or less toxic than the original contaminant. Many contaminants adsorb to organic matter, which can make them less bioavailable.

Environmental factors that influence the bioavailability of a contaminant in water are important to aquatic risk assessments. Factors including pH, hardness, or aerobic status can determine both the chemical form and uptake of contaminants by biota. Other environmental factors can influence how organisms process contaminants. For example, as water temperatures rise, metabolism of fish and aquatic invertebrates increases, and the rate of uptake of a contaminant from water can increase.

If the literature search on the contaminants of concern reveals information on the bioavailability of a contaminant, then appropriate bioaccumulation or bioconcentration factors (BAFs or BCFs) for the contaminants should be determined. If not readily available in the literature, BAF or BCF values can be estimated from studies that report contaminant concentrations in both the environmental exposure medium (e.g., sediments) and in the exposed biota (e.g., benthic macroinvertebrates). Caution is necessary, however, when extrapolating BAF or BCF values estimated for one ecosystem to another ecosystem.

Life history. Because it is impossible and unnecessary to model an entire ecosystem, the selection of assessment endpoints and associated species of concern, and measurement endpoints (including those for a surrogate species if necessary) are fundamental to a successful risk assessment. This process is described in Steps 3 and 4. Once assessment and measurement endpoints are agreed to by the risk assessor and risk manager, life history information for the species of concern or the surrogate species should be collected. Patterns of activity and feeding habits of a species affect their potential for exposure to a contaminant (e.g., grooming activities of small mammals, egestion of bone and hide by owls). Other important exposure factors include food and water ingestion rates, composition of the diet, average body weight, home range size, and seasonal activities such as migration.

Ecological effects profile. Once contaminants and species of concern are selected during problem formulation, a general overview of toxicity and toxic mechanisms is needed. The distinction between the species of concern representing an assessment endpoint and a surrogate species representing a measurement endpoint is important. The species of concern is the species that might be threatened by contaminants at the site. A surrogate species is used when it is not appropriate or possible to measure attributes of the species of concern. A surrogate for a species of concern should be sufficiently similar biologically to allow inferences on likely effects in the species of concern.

The ecological effects profile should include toxicity information from the literature for each possible exposure route. A lowest-observed-adverse-effect level (LOAEL) and the no-observed-adverse-effect level (NOAEL) for the species of concern or its surrogate should be obtained. Unfortunately, LOAELs are available for few wildlife species and contaminants. If used with caution, toxicity data from a closely related species can be used to estimate a LOAEL and a NOAEL for a receptor species.

C-2 INFORMATION SOURCES

This section describes information sources that can be examined to find the information described in Section 3-1. A logical and focused literature search will reduce the time spent searching for pertinent information.

A first step in a literature search is to develop a search strategy, including a list of key words. The next step is to review computerized databases, either on-line or CD-ROM-based information systems. These systems can be searched based on a number of parameters.

Scientific abstracts that contain up-to-date listings of current, published information also are useful information sources. Most abstracts are indexed by author or subject. Toxicity studies and information on wildlife life-histories often are summarized in literature reviews published in books or peer-reviewed journals. Original reports of toxicity studies can be identified in the literature section of published documents. The original article in which data are reported must be reviewed before the data are cited in a risk assessment.

Key words. Once the risk assessor has prepared a list of the specific information needed for the risk assessment, a list of key words can be developed. Card catalogs, abstracts, on-line databases, and other reference materials usually are indexed on a limited set of key words. Therefore, the key words used to search for information must be considered carefully.

Useful key words include the contaminant of concern, the biological species of concern, the type of toxicity information wanted, or other associated words. In addition, related subjects can be used as key words. However, it usually is necessary to limit peripheral aspects of the subject in order to narrow the search. For example, if the risk assessor needs information on the toxicity of lead in soils to moles, then requiring that both "lead" and "mole" are among the key words can focus the literature search. If the risk assessor needs information on a given plant or animal species (or group of species), key words should include both the scientific name (e.g., genus and species names or order or family names) and an accepted common name(s). The projected use of the data in the risk assessment helps determine which key words are most appropriate.

If someone outside of the risk assessment team will conduct the literature search, it is important that they understand both the key words and the study objectives for the data.

Databases. Databases are usually on-line or CD-ROM-based information systems. These systems can be searched using a number of parameters. Prior to searching databases, the risk assessor should determine which database(s) is most likely to provide the information needed for the risk assessment. For example, U.S. Environmental Protection Agency's (EPA's) AQUIRE database (AQUatic Information REtrieval database) provides information specifically on the toxicity of chemicals to aquatic plants and animals. PHYTOTOX includes data on the toxicity of contaminants to terrestrial and aquatic plants, and TERRETOX includes data on toxicity to terrestrial animals. U.S. EPA's IRIS (Integrated Risk Information System) provides information on human health risks (e.g., references to original toxicity studies) and regulatory information (e.g., reference doses and cancer potency factors) for a variety of chemicals. Other useful databases include the National Library of Medicine's HSDB (Hazardous Substances Data Bank) and the National Center for Environmental Assessment's HEAST Tables (Health Effects Assessment Summary Tables). Commercially available databases include BIOSIS (Biosciences Information Services) and ENVIROLINE. Another database, the U.S. Public Health Service's Registry of Toxic Effects of Chemical Substances (RTECS) is a compilation of toxicity data extracted from the scientific literature and is also available online.

Several states have *Fish and Wildlife History Databases* or *Academy of Science* databases, which often provide useful information on the life-histories of plants and animals in the state. State databases are particularly useful for obtaining information on endemic organisms or geographically distinct habitats.

Databases searches can yield a large amount of information in a short period of time. Thus, if the key words do not accurately describe the information needed, database searches can provide a large amount of irrelevant information. Access fees and on-line fees can apply; therefore, the selection of relevant key words and an organized approach to the search will reduce the time and expense of on-line literature searches.

Abstracts. Published abstract compilations (e.g., Biological Abstracts, Chemical Abstracts, Applied Ecology Abstracts) contain up-to-date listings of current, published information. Most abstracts are indexed by author or subject. Authors and key words can be cross-referenced to identify additional publications. Abstract compilations also include, for each citation, a copy of its abstract from the journal or book in which it was published. Reviewing the abstracts of individual citations is a relatively quick way to determine whether an article is applicable to the risk assessment. As with computerized database searches, it is important to determine which abstract compilations are most suitable for the risk assessor's information needs.

Published abstract compilations that are indexed by author are particularly useful. If an author is known to conduct a specific type of research, their name would be referenced in the abstract for other articles on similar subjects. If the risk assessor considers an abstract pertinent to the assessment, the original article must be retrieved and reviewed before it can be cited in the risk assessment. Otherwise, the results of the risk assessment could be based on incorrect and incomplete information about a study.

Abstracts usually must be searched manually, which can be a very time consuming. The judicious use of key words can help to reduce the amount of time needed to search through these volumes.

Literature review publications. Published literature reviews often cover toxicity or wildlife information of value to an ecological risk assessment. For example, the U.S. Fish and Wildlife Services (U.S. FWS) has published several contaminant-specific documents that list toxicological data on terrestrial, aquatic, and avian studies (e.g., Eisler, 1988). The U.S. EPA publishes ambient water quality criteria documents (e.g., U.S. EPA, 1985) that list all the data used to calculate those values. Some literature reviews critically evaluate the original studies (e.g., toxicity data reviewed by NOAA, 1990). The *Wildlife Exposure Factors Handbook* (U.S. EPA, 1993a,b) provides pertinent information on exposure factors (e.g., body weights, food ingestion rates, dietary composition, home range size) for 34 selected wildlife species.

Literature reviews can provide an extensive amount of information. However, the risk assessor must obtain a copy of the original of any studies identified in a literature review that will be used in the risk assessment. The original study must be reviewed and evaluated before it can be used in the risk assessment. Otherwise, the results of the risk assessment could be based on incorrect and incomplete information about a study.

References cited in previous studies. Pertinent studies can be identified in the literature cited section of published documents that are relevant to the risk assessment, and one often can identify several investigators who work on related studies. Searching for references in the literature cited section of published documents, however, takes time and might not be very effective. However, this is probably the most common approach to identifying relevant literature. If this approach is selected, the best place to start is a review article. Many journals do not list the title of a citation for an article, however, limiting the usefulness of this technique. Also, it can be difficult to retrieve literature cited in obscure or foreign journals or in unpublished masters' theses or doctoral dissertations. Although this approach tends to be more time consuming than the other literature search approaches described above, it probably is the most common approach used to locate information for a risk assessment.

Journal articles, books, government documents. There are a variety of journals, books, and government documents that contain information useful to risk assessments. The same requirement for retrieving the original reports for any information used in the risk assessment described for other information sources applies to these sources.

Threatened and endangered species. Threatened and endangered species are of concern to both federal and state governments. When conducting an ecological risk assessment, it often is necessary to determine or estimate the effects of site contaminants to federal threatened or endangered species. In addition, other special-status species (e.g., species listed by a state as endangered or threatened within the state) also can be the focus of the assessment. During the problem formulation step, the U.S. FWS or state Natural Heritage programs should be contacted to determine if these species are present or might be present on or near a Superfund site.

Once the presence of a special-status species is confirmed or considered likely, information on this species, as well as on surrogate species, should be included in the literature search. There are specific federal and state programs that deal with issues related to special-status species, and often there is more information available for these than for non-special-status species used as surrogates for an ecological risk assessment. Nonetheless, the use of surrogate species usually is necessary when an assessment endpoint is a special-status species.

REFERENCES

Eisler, R. 1988. Lead Hazards to Fish, Wildlife, and Invertebrates: A Synoptic Review. U.S. Fish and Wildlife Service Patuxent Wildlife Research Center, Laurel MD: U.S. Department of the Interior; Biological Report 85(1.14), Contaminant Hazard Reviews Rep. No. 14.

National Oceanic and Atmospheric Administration (NOAA). 1990. *The Potential for Biological Effects of Sediment-Sorbed Contaminants Tested in the National Status and Trends Program*. Seattle, WA: Office of Oceanography and Marine Assessment. NOAA/TM/NOS/OMA-52. Technical memorandum by Long, E.R. and Morgan, L.G.

U.S. Environmental Protection Agency (U.S. EPA). 1993a. *Wildlife Exposure Factors Handbook Volume I*. Washington, DC: Office of Research and Development; EPA/600/R-93/187a.

U.S. Environmental Protection Agency (U.S. EPA). 1993b. *Wildlife Exposure Factors Handbook Volume II: Appendix*. Washington, DC: Office of Research and Development; EPA/600/R-93/187b.

U.S. Environmental Protection Agency (U.S. EPA). 1985. *Ambient Water Quality Criteria for Copper-1984*. Washington, DC: Office of Water, Regulations and Standards, Criteria and Standards Division. EPA/440/5-84-031. PB85-227023.

APPENDIX D
STATISTICAL CONSIDERATIONS

APPENDIX D
STATISTICAL CONSIDERATIONS

In the biological sciences, statistical tests often are needed to support decisions based on alternative hypotheses because of the natural variability in the systems under investigation. A statistical test examines a set of sample data, and, based on an expected distribution of the data, leads to a decision on whether to accept the hypothesis underlying the expected distribution or whether to reject that hypothesis and accept an alternative one. The null hypothesis is a hypothesis of no differences. It usually is formulated for the express purpose of being rejected. The alternative or test hypothesis is an operational statement of the investigator's research hypothesis. An example of a null hypothesis for toxicity testing would be that mortality of water fleas exposed to water from a contaminated area is no different than mortality of water fleas exposed to water from an otherwise similar, but uncontaminated area. An example of the test hypothesis is that mortality of water fleas exposed to water from the contaminated area is higher than mortality of water fleas exposed to uncontaminated water.

D-1 TYPE I AND TYPE II ERROR

There are two types of correct decisions for hypothesis testing: (1) accepting a true null hypothesis, and (2) rejecting a false null hypothesis. There also are two types of incorrect decisions: rejecting a true null hypothesis, called Type I error; and accepting a false null hypothesis, called Type II error.

When designing a test of a hypothesis, one should decide what magnitude of Type I error (rejection of a true null hypothesis) is acceptable. Even when sampling from a population of known parameters, there are always some sample sets which, by chance, differ markedly. If one allows 5 percent of samples to lead to a Type I error, then one would on average reject a true null hypothesis for 5 out of every 100 samples taken. In other words, we would be confident that, 95 times out of 100, one would not reject the null hypothesis of no difference "by mistake" (because chance alone produced such deviant results). When the probability of Type I error (commonly symbolized by α) is set at 0.05, this is called a significance level of 5 percent. Setting a significance level of 5 percent is a widely accepted convention in most experimental sciences, but it is just that, a convention. One can demand more confidence (e.g., $\alpha = 0.01$) or less confidence (e.g., $\alpha = 0.10$) that the hypothesis of no difference is not rejected by mistake.

If one requires more confidence for a given sample size that the null hypothesis is not rejected by mistake (e.g., $\alpha = 0.01$), the chances of Type II error increase. In other words, the chance increases that one will mistakenly accept a false null hypothesis (e.g., mistakenly believe that the contaminated water from the site has no effect on mortality of water fleas). The probability of Type II error is commonly denoted by β.

Thus:

p (Type I error) = α

p (Type II error) = β

However, if one tries to evaluate the probability of Type II error (accepting a false hypothesis of no difference), there is a problem. If the null hypothesis is false, then some other hypothesis must be true, but unless one can specify a second hypothesis, one can't determine the probability of Type II error. This leads to another important statistical consideration, which is the power of a study design and the statistical test used to evaluate the results.

D-2 STATISTICAL POWER

The power of a statistical test is equal to (1 - β) and is equal to the probability of rejecting the null hypothesis (no difference) when it should be rejected (i.e., it is false) and the specified alternative hypothesis is true. Obviously, for any given test (e.g., a toxicity test at a Superfund site), one would like the quantity (1 - β) to be as large as possible (and β to be as small as possible). Because one generally cannot specify a given alternative hypothesis (e.g., mortality should be 40 percent in the exposed population), the power of a test is generally evaluated on the basis of a continuum of possible alternative hypotheses.

Ideally, one would specify both α and β before an experiment or test of the hypothesis is conducted. In practice, it is usual to specify α (e.g., 0.05) and the sample size because the exact alternative hypothesis cannot be specified.[1] Given the inverse relationship between the likelihood of making Type I and Type II errors, a decrease in α will increase β for any given sample size.

To improve the statistical power of a test (i.e., reduce β), while keeping α constant, one can either increase the sample size (N) or change the nature of the statistical test. Some statistical tests are more powerful than others, but it is important that the assumptions required by the test (e.g., normality of the underlying distribution) are met for the test results to be valid. In general, the more powerful tests rely on more assumptions about the data (see Section D-3).

Alternative study designs sometimes can improve statistical power (e.g., stratified random sampling compared with random sampling if something is known about the history and location of contaminant release). A discussion of different statistical sampling designs is beyond the scope of this guidance, however. Several references provide guidance on statistical sampling design, sampling techniques, and statistical analyses appropriate for hazardous waste sites (e.g., see Cochran, 1977; Green, 1979; Gilbert, 1987; Ott, 1995).

One also can improve the power of a statistical test if the test hypothesis is more specific than "two populations are different," and, instead, predicts the direction of a difference (e.g., mortality in

---

[1] With a specified alternative hypothesis, once α and the sample size (N) are set, β is determined.

the exposed group is higher than mortality in the control group). When one can predict the direction of a difference between groups, one uses a one-tailed statistical test; otherwise, one must use the less powerful two-tailed version of the test.

Highlight D-2
Key Points About Statistical Significance, Power, and Sample Size

(1) The significance level for a statistical test, α, is the probability that a statistical test will yield a value under which the null hypothesis will be rejected when it is in fact true. In other words, α defines the probability of committing Type I error (e.g., concluding that the site medium is toxic when it is in fact not toxic to the test organisms).

(2) The value of β is the probability that a statistical test will yield a value under which the null hypothesis is accepted when it is in fact false. Thus, β defines the probability of committing Type II error (e.g., concluding that the site medium is not toxic when it is in fact toxic to the test organisms).

(3) The power of a statistical test (i.e., 1 - β) indicates the probability of rejecting the null hypotheses when it is false (and therefore should be rejected). Thus, one wants the power of a statistical test to be as high as possible.

(4) Power is related to the nature of the statistical test chosen. A one-tailed test is more powerful than a two-tailed test. If the alternative to the null hypothesis can state the expected direction of a difference between a test and control group, one can use the more powerful one-tailed test.

(5) The power of any statistical test increases with increasing sample size.

D-3 STATISTICAL MODEL

Associated with every statistical test is a model and a measurement requirement. Each statistical test is valid only under certain conditions. Sometimes, it is possible to test whether the conditions of a particular statistical model are met, but more often, one has to assume that they are or are not met based on an understanding of the underlying population and sampling design. The conditions that must be met for a statistical test to be valid often are referred to as the assumptions of the test.

The most powerful statistical tests (see previous section) are those with the most extensive assumptions. In general, parametric statistical tests (e.g., t test, F test) are the most powerful tests, but also have the most exacting assumptions to be met:

(1) The "observations" must be independent;

(2) The "observations" must be drawn from a population that is normally distributed;

(3) The populations must have the same variance (or in special cases, a known ratio of variances); and

(4) The variables must have been measured at least on an interval scale so that it is possible to use arithmetic operations (e.g., addition, multiplication) on the measured values (Siegel, 1956).

The second and third assumptions are the ones most often violated by the types of data associated with biological hypothesis testing. Often, distributions are positively skewed (i.e., longer upper than lower tail of the distribution). Sometimes, it is possible to transform data from positively skewed distributions to normal distributions using a mathematical function. For example, many biological parameters turn out to be log-normally distributed (i.e., if one takes the log of all measures, the resulting values are normally distributed). Sometimes, however, the underlying shape of the distribution cannot be normalized (e.g., it is bimodal).

When the assumptions required for parametric tests are not met, one must use nonparametric statistics (e.g., median test, chi-squared test). Nonparametric tests are in general less powerful than parametric tests because less is known or assumed about the shape of the underlying distributions. However, the loss in power can be compensated for by an increase in sample size, which is the concept behind measures of power-efficiency.

Power-efficiency reflects the increase in sample size necessary to make test B (e.g., a nonparametric test) as efficient or powerful as test A (e.g., a parametric test). A power-efficiency of 80 percent means that in order for test B to be as powerful as test A, one must make 10 observations for test B for every 8 observations for test A.

For further information on statistical tests, consult references on the topic (e.g., see references below).

REFERENCES

Cochran, W. G. 1977. *Sampling Techniques*. *Third edition*. New York, NY: John Wiley and Sons, Inc.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring*. New York, NY: Reinhold.

Green, R. H. 1979. *Sampling Design and Statistical Methods for Environmental Biologists*. New York, NY: Wiley.

Ott, W.R. 1995. *Environmental Statistics and Data Analysis*. Boca Raton, FL: CRC Press, Inc., Lewis Publishers.

Siegel, S. 1956. *Non-parametric Statistics*. New York, NY: McGraw-Hill.