
Prepared for
General Electric Company

Prepared by:
ARCADIS Geraghty & Miller, Inc.
Portland, Maine
Tel 207 828 0046

JSA Environmental
Long Beach, California
Tel 562 628 1175

Exponent
Bellevue, Washington
Tel 425 643 9803

Date
June 3, 1999

This document is intended only for the use of the individual or entity for which it was prepared and may contain information that is privileged, confidential, and exempt from disclosure under applicable law. Any dissemination, distribution, or copying of this document is strictly prohibited.
June 7, 1999

Mr. Richard Cavagnero  
GE Project Leader  
U.S. Environmental Protection Agency  
Office of Site Remediation and Restoration  
One Congress Street, Suite 1100  
Boston, MA 02114

Re: Comments on EPA's Supplemental Investigation Work Plan for the Lower Housatonic River

Dear Mr. Cavagnero:

As we have expressed to EPA on several prior occasions, GE has substantial concerns about the extent of sampling and studies that EPA is carrying out and planning to carry out on the "Rest of River" portion of the Pittsfield/Housatonic River Site. We are herewith providing detailed written comments on EPA's Work Plan for such sampling and studies, entitled Supplemental Investigation Work Plan for the Lower Housatonic River (Roy F. Weston, Inc., February 10, 1999). These comments consist of four separate sets of comments:

1. Comments on USEPA Supplemental Investigation Work Plan for Lower Housatonic River Relating to Fate, Transport and Bioaccumulation Modeling, prepared on GE's behalf by Quantitative Environmental Analysis, LLC;

2. Comments of the General Electric Company on EPA's Human Health Risk Assessment Work Plan for the Lower Housatonic River, prepared on GE's behalf by Ogden Environmental and Energy Services;

3. Comments of the General Electric Company on EPA's Ecological Risk Assessment Work Plan for the Lower Housatonic River, prepared on GE's behalf by ARCADIS Geraghty and Miller, JSA Environmental, and Exponent; and

4. Miscellaneous Comments on Work Plan, prepared on GE's behalf by Blasland, Bouck & Lee, relating to portions of the Work Plan that do not pertain specifically to the modeling and risk assessment activities.
We look forward to discussing these comments with EPA.

Sincerely,

Andrew T. Silfer, P.E.
Remediation Project Manager

Enclosures

cc:  Tim Conway, EPA
     John Kilborn, EPA*
     Bryan Olson, EPA*
     Susan Svirsky, EPA*
     Dean Tagliaferro, EPA*
     Steve Botts, EPA
     Amy Legare, EPA
     Deborah Jones, WESTON*
     Tod DeLong, Avatar Environmental*
     Cynthia Huber, DOJ
     Addie Fiske, DOJ
     Ralph Child, MDEP
     Alan Weinberg, MDEP*
     Robert Bell, MDEP
     Matthew Brock, MA AG
     Betsy Harper, MA AG
     Tom O'Brien, MA EOEA
     Ken Finkelstein, NOAA*
     Anton Giedt, NOAA
     Richard Webb, CT AG
     Steve Ramsey, GE
     Jane Gardner, GE
     Kevin Holtzclaw, GE*
     Andrew Thomas, GE*
     Leonard Shen, GE
     Michael Carroll, GE*
     James Bieke, Shea & Gardner*
     David Buente, Sidley & Austin*
     Robert Goldman, BBL*
John Connolly, QEA*
Russell Keenan, Ogden*
Kenneth Jenkins, JSA*
Thomas Ginn, Exponent*
Miranda Henning, ARCADIS Geraghty & Miller*

(* with enclosures)
TO: Recipients of GE Comments on the Supplemental Investigation Work Plan, Ecological Risk Assessment

FROM: Ellen Losano-Ramsey

PROJECT: GE Housatonic River Project

DATE: 18 June 1999

W. O. NO.: 10971.032.002.0149

SUBJECT: Comments from GE: Cover Letter and Ecological Risk Assessment Work Plan for the Lower Housatonic River

ACTION:

Attached are the cover letter and GE comments on the Ecological Risk Assessment Work Plan for the Lower Housatonic River. These replace the previously sent versions, which have been withdrawn. Please discard the earlier versions and use this information.

If you have any questions, please contact me at 610-701-3078.
1.0 INTRODUCTION

Exponent, ARCADIS Geraghty & Miller, and the latter's subsidiary JSA Environmental were retained by General Electric Company (GE) to review and provide comments on the Supplemental Investigation Work Plan for the Lower Housatonic River prepared by WESTON (1999) on behalf of the U.S. Environmental Protection Agency (EPA) and the U.S. Army Corps of Engineers (ACE). These comments are provided on the work plan with the intention of guiding the conduct of the ecological risk assessment (ERA) early in the process, to ensure that the data collected will be meaningful and useful for risk management decisions. The following comments reflect discussions held with EPA and its contractors on May 6, 1999, and in a subsequent conference call on May 14, 1999 which focused on the benthic invertebrate and sediment toxicity studies. It should be noted that the discussion that follows addresses only those components of the work plan that relate to the ERA for the Lower Housatonic River. Comments on other components of the work plan (including biological sampling for the bioaccumulation model that is part of the overall polychlorinated biphenyl (PCB) fate, transport, and bioaccumulation model for the Lower Housatonic River) will be submitted under separate cover.

A number of significant concerns are common to most, if not all, components of the work plan. These general comments are listed below, and then explained in detail in Section 2.0. Section 3.0 provides specific comments on individual sections of the work plan.

1. Because a number of the studies described in the work plan have already commenced or been carried out, GE (and others) are precluded from a meaningful opportunity to discuss those studies with EPA.
2. The work plan is inconsistent, ambiguous, and incomplete.
3. The approach to risk characterization is vague and lacks an objective method for integrating multiple lines of evidence.
4. Most of the assessment and measurement endpoints identified in the work plan's problem formulation focus on individuals, rather than subpopulations, populations, or communities.
5. Several proposed studies have no risk-related objectives.
6. Nature and intensity of effects, spatial and temporal scale, and potential for recovery are inadequately addressed in the proposed study designs.

7. The work plan relies too much on comparisons to literature-derived maximum acceptable tissue concentrations (MATCs) and reference toxicity values (RTVs), in place of site-specific data.

8. The work plan inappropriately proposes to rely on generic benchmarks, such as sediment quality guidelines (SQG), which may be useful tools for screening assessments, but are not appropriate for a baseline risk assessment conducted at this scale.

9. Parameters to be used in food chain modeling to estimate dose are unreasonable and unrealistic, or are not specified.

10. The majority of proposed field studies do not specify criteria used to select reference areas, the locations of candidate reference areas, or whether reference areas will even be evaluated. For those studies that do acknowledge that reference areas will be evaluated, the number of reference areas is not sufficient to characterize natural variability in the parameters under evaluation.

11. Some of the in situ toxicity tests are inappropriate because they are not interpretable and do not provide additional useful data to other proposed studies.

12. Although Section 7 of the work plan repeatedly states that an effort will be made to define no-effects levels and stressor-response curves as part of the individual studies, many of those study designs are not adequate to meet that objective.

13. The proposed methods for exposure assessment do not address and/or meet the minimal requirements for spatial data and, therefore, do not enable spatially explicit decisions to be made during risk management.

14. There is no analytical or interpretative framework for use of total PCB or Aroclor data vs. PCB congener-specific data, although both types of data will be collected.

15. Many of the non-standardized test protocols and field work appear to be research oriented and not directly applicable to the estimation of ecological risks and risk management.

16. The extent of the sediment sampling and chemical analysis programs proposed appears to be out of proportion with most of the ecological studies proposed.

17. The protocols presented in the appendices lack adequate detail and/or are poorly designed.

Examples, implications, and recommendations related to these 17 general problems are detailed in Section 2.0. Specific comments are presented in Section 3.0, including numerous examples of the 17 general problems.
2.0 GENERAL COMMENTS

The following discussions further describe the 17 general points listed in the introduction providing examples, implications, and recommendations.

2.1 A number of the studies described have already commenced or been carried out.

The following activities appear to have been completed in 1998: collection and analysis of over 1,000 fish samples, waterfowl collection and analysis, surveys of forest birds and marsh and wading birds, and the first phases of the tree swallow reproduction study and the reptile and amphibian use survey. In addition, the work plan was issued only a few months before the start of the 1999 field season, when all of the remaining studies will be conducted. Consequently, GE (and others) have been precluded from having a meaningful opportunity to discuss those studies with EPA prior to their implementation. As detailed in Section 3.0, virtually all of the proposed studies have important limitations that will substantially affect their usefulness and interpretability for risk assessment purposes. Hence, both the protocols for those studies and the field work will likely need to be redone, or the studies should be given little or no weight in the ERA's risk characterization.

2.2 The work plan is inconsistent, ambiguous, and incomplete.

Numerous key elements are missing from the work plan. For example, the work plan fails to explain how each study will be used in the ERA (e.g., within the problem formulation to describe the fauna at the site, to assist in the selection of receptors or the design of other studies, or as a measurement endpoint). The work plan also does not provide rationale or criteria for selecting individual species for evaluation, or locations or numbers of study areas and reference areas. The vast majority of appendices omit any discussion of the statistical design of the studies or the basis for the specified sample sizes. Model formulations for the prediction of water and sediment quality, as well as food chain transfer of PCBs, are not described; consequently, the relationship of some field data to the models is unclear. The study designs do not provide criteria that will be applied in interpreting the results of the studies. Furthermore, study designs are often internally inconsistent (e.g., tree swallows, fish, mussels). As an example, the second paragraph of the fish protocol (App. A.10, p. 3) refers to five study areas (line 5), while later in that paragraph and in Table 1 only four study areas are
discussed. Discussions within Sections 5 and 7 and the appendices that relate to the numbers of samples to be collected are often inconsistent. Consequently, it is often difficult to determine the scope and goals of the individual studies, the specific methods to be employed, and the risk questions to be answered. The work plan should be substantially revised to correct all of these problems.

2.3 *The approach to risk characterization is vague and lacks an objective method to integrate multiple lines of evidence.*

Although the work plan indicates that the qualitative weight of evidence approach developed by Menzie et al. (1996) will be employed during problem formulation (p. 7-15), the procedure apparently was not conducted prior to selecting measurement endpoints. Since the principal purposes of the weight of evidence approach are to ensure the scientific defensibility of measurement endpoints and interpretation of potential risks, and to bring transparency to the planning and risk characterization processes, the work plan’s problem formulation should walk through this procedure prior to selecting measurement endpoints. Additionally, several other approaches for analyzing data are mentioned in the work plan’s discussion of risk characterization, but the approach to be taken is not identified. Indeed, the weight of evidence approach discussed during problem formulation is not revisited in risk characterization. As such, there is no linkage of the risk characterization approach to assessment endpoints. Likewise, the possible use of probabilistic methods is mentioned, but no details are provided (e.g., which receptors and models will undergo probabilistic analysis, distributions that will be applied). The work plan should be revised to specify a transparent, objective, and scientifically defensible basis for risk characterization, in order to ensure that bias is not introduced in the selection of endpoints and interpretation of analytical results.

2.4 *Most of the assessment and measurement endpoints identified in the work plan focus on individuals, rather than subpopulations, populations, or communities (which are the proper focus of ERA, particularly over such a large spatial and seasonal scale).*

EPA’s (1998a) draft *Risk Management Guidance* clearly states that populations are the appropriate level of ecological organization for assessment. Similarly, the Massachusetts Contingency Plan (MCP) provides that “a level of no significant risk of harm to the environment exists, or has been achieved, if... there is no evidence of biologically significant harm (at the *subpopulation, community or system-wide level*) known or believed to be associated with current or foreseeable future exposure of wildlife, fish, shellfish, or other aquatic biota to oil and/or hazardous material at or from the
disposal site...” (emphasis added) (310 CMR 40.0995(4)(d)). However, with few exceptions, survival, growth and reproduction of individuals are the assessment endpoints identified for virtually all receptors. Additionally, existing data on populations and communities (e.g., great blue heron reproduction, passerine reproduction and community structure, fish communities) are not considered, and no rationale for excluding such studies is provided. The few field surveys proposed that do consider community-level endpoints (e.g., species richness) for amphibians, reptiles, and birds are not included among the assessment and measurement endpoints listed in Tables 4.2-1 and 7-1. Reliance on assessment endpoints that focus on individuals is not useful from a risk management standpoint, and will yield substantial overestimates of potential risks to higher levels of ecological organization. The work plan should be revised to place greater emphasis on measurement endpoints focused at higher levels of ecological organization. For example, more rigorous survey designs for birds and mammals would allow evaluation of community structure as a measurement endpoint. Likewise, existing data on great blue heron reproduction, passerine reproduction and community structure, and fish community structure should be considered for inclusion as measurement endpoints.

2.5 Several proposed studies have no risk-related objectives.

For example, several surveys (e.g., rare plants and communities, dragonflies, birds, mammals) appear to gather only descriptive data, rather than information that can be related to prediction of potential site-related risks. Other studies (e.g., earthworms, aquatic invertebrates, crayfish, waterfowl, small mammals) focus on tissue sampling and analysis, thereby treating bioaccumulation as an endpoint in itself, rather than ascertaining whether the populations are actually adversely affected. In order for any study to yield results that will be meaningful in ERA, the findings must include measures of effects that can be compared either: (a) to parallel measurements made at reference areas; (b) across a concentration gradient; or (c) to literature reference/benchmark values. The study designs listed above should be revised so that they will discern differences between the site and the points of comparison or along a chemical concentration gradient, thereby addressing risk-related objectives. Furthermore, the work plan should specify clear links between literature-based endpoints focused on individual organisms (e.g., toxicity quotients, or TQs) and site-specific surveys focused on populations and/or communities.
2.6 There is no evidence that (a) nature and intensity of effects, (b) spatial and temporal scale, and (c) potential for recovery were considered in the design of the proposed studies.

Section 7.4.3.2 provides a general discussion of how the "level of significance and risk to any adverse responses" will be considered in relation to nature and intensity of effects, spatial and temporal scale and potential for recovery. However, it does not appear that these factors were adequately considered in the design of studies presented in the work plan. Similarly, there is no discussion of how the assessment and measurement endpoints presented in Table 7-1 will be used to assess these concepts. In fact, several of the study designs directly conflict with these objectives. For example, the designs for the mussel, earthworm and fish studies do not provide a basis for evaluating normal temporal variability. Additionally, for most proposed studies (including amphibians, reptiles, birds, and mammals), reference locations are inadequate (and/or inadequately defined) to provide the basis for establishing normal ecosystem variability. While some study designs (e.g. tree swallow and small mammal studies) propose to collect many more samples than are necessary for characterization purposes, for other study designs, the proposed extent of sampling is insufficient to establish an adequate duration or frequency of exposure (e.g., algal, phytoplankton, zooplankton, and macrophyte bioaccumulation). Finally, none of the study designs address the potential for recovery. As a result, the studies described in this work plan are inadequate to provide a basis for evaluating the ecological significance and risk of observed or predicted adverse responses, which will likely lead to a subjective evaluation of ecological significance.

2.7 The work plan places too much reliance on predictive methods, such as comparisons to literature-derived MATCs and RTVs, in place of site-specific information.

Predictive techniques generally involve the use of food web modeling to estimate the concentrations of the chemicals of potential concern (COPCs) in the diet or tissue of the receptor of interest, based on concentrations in the sediment, soil, water, or prey. The estimated intake or tissue concentrations are then compared with RTVs or MATCs derived from the scientific literature — i.e., values that have been found to be unassociated with adverse effects — in order to yield a quotient indicative of the potential for adverse effects in the receptor. Under the work plan, MATCs will be used in the evaluation of earthworms, fish, waterfowl, and small mammals, while evaluations of other birds and mammals will depend on RTVs. Such predictive techniques are useful screening tools for assessing the potential for ecological risks, particularly for receptors for which field studies are not feasible or are not likely to be sufficiently sensitive to detect adverse effects.
However, predictive techniques also have significant uncertainties and limitations. First, MATCs and RTVs rarely incorporate information from the entire stressor-response relationship, because they are usually limited to single point comparisons [e.g., no-observed-effect levels (NOEL) or lowest observed effects levels (LOEL)]. Second, where site-specific field data exist or are proposed to be collected, generic benchmark comparisons are unwarranted and only dilute interpretation of the results.

Third, given that predictive techniques generally rely on the scientific literature for information regarding exposure and toxicity, the certainty of the results is directly dependent on the quality of available information. For many species and many chemicals, very limited relevant information has been published. As a result, it is often necessary to extrapolate from the available literature on related species and chemicals. The uncertainty factors applied to such extrapolations, to account for interspecies differences in sensitivity and exposure potential, may be arbitrarily set at factors of ten, resulting in overly conservative estimates of risk.

Fourth, predictive techniques are rarely adjusted to account for potential effects at the population or community level; because it is most straightforward to calculate hazards to individual organisms, predictive approaches generally make the highly conservative assumption that effects to individual organisms are predictive of effects to more complex levels of ecological organization. In reality, numerous site-specific factors may mitigate exposure, and density-dependent mechanisms can allow natural populations to compensate for perturbations in their size and age structures; predictive techniques cannot adequately account for such natural factors and mechanisms.

The above limitations of predictive methods are recognized by regulatory agencies. The Massachusetts Department of Environmental Protection (DEP) (1996) notes that “[b]ased on a comparison of the results of predictive methods with direct measurement, investigators have reported that predictive methods may substantially overestimate risk for organisms at higher trophic levels.” EPA (1998b) also describes numerous limitations of such predictive approaches and recommends that the output of models be compared with actual measurements in the system of interest.

Given these limitations, GE recommends consideration of a phased approach to predictive techniques, as follows. As a first step, a relatively simple TQ would be calculated as a conservative screen, based on the ratio of the estimated intake of a chemical by a given individual receptor to an RTV derived or
extrapolated from the literature. If this conservative screen shows no effects, then there would be no need for further study. However, if the TQ screening procedure predicts possible adverse effects on the individual organism, that prediction is not sufficient to demonstrate a significant risk at the population, community, or ecosystem level. In such cases, further work should be conducted, in order to incorporate more site-specific and more complete information into the TQ approach and/or to determine whether the predicted effects on the individual organism have adverse impacts at the population or community level. Such additional work may include: (a) an effort to obtain more site-specific estimates of exposure for input into the TQ estimate; (b) a probabilistic TQ analysis using distributions of input parameter values (instead of single-point values); (c) modifications of the RTV so that it reflects more complete information (e.g., benchmark dose approach, use of geometric mean of appropriate values); (d) population-level modeling appropriate for the receptors of concern; and/or (e) collections of field data to determine whether the predicted effects are verified in the field.

Additionally, because the TQ approach requires comparison of like measurements ("apples to apples"), the receptors of interest, RTVs and MATCs should be defined in advance of collecting and analyzing soil, sediment, surface water, or prey samples. Doing so will help ensure that: (a) appropriate prey species and sizes are targeted; (b) sample preparation parallels that associated with the MATC; and (c) analytical methods are comparable.

2.8 The work plan inappropriately proposes to rely on generic sediment quality and water quality benchmarks, which may be useful tools for screening level assessments, but are not appropriate for a baseline ERA conducted on this scale.

Chemical concentrations in sediment will be compared to several different SQGs, despite the proposal to evaluate the actual benthic invertebrate community structure at the site. The use of generic benchmarks as lines of evidence in a baseline ERA is inappropriate for a number of reasons. First, generic benchmarks do not account for site-specific conditions. Second, because the proposed SQGs are based on sediments contaminated with multiple chemicals, it is not possible to attribute toxicity to individual chemicals. Third, the results of the assessment will not be scientifically defensible because some of the benchmarks proposed are technically flawed [e.g., Ontario Ministry of the Environment's Lowest Effect Levels (LELs) and Severe Effect Levels (SELs) and National Oceanographic and Atmospheric Administration's (NOAA) Effects Range Low (ER-Ls) and Effects Range Median (ER-Ms)]. For example, the method used to establish ER-L and ER-M values relies only on the distribution of effects data and fails to include the information conveyed by samples in which no toxic
effects are observed. Although this is a reasonable approach when evaluating spiked sediment bioassay data, it is an inappropriate approach when applied to naturally occurring sediments, like the Housatonic River, in which the compound(s) responsible for observed toxicity are unknown and possibly unmeasured. The consequence is that ER-M and ER-L values are not reliable indicators of organism responses to chemicals in sediment. Attempts to field validate the ER-L and ER-M data have demonstrated that the error rates for false positives are unacceptably high. For organic compounds, many of the false positive rates for ER-Ls and ER-Ms are >25 percent. Because of these uncertainties, Long et al. (1995) caution that “[t]he numerical guidelines should be used as informal screening tools in environmental assessments.” Another NOAA scientist has concluded that: “exceedance of any particular SQG means very little with regard to sediment toxicity” (O’Connor 1999).

As with the use of MATCs and RTVs, the use of generic benchmarks as lines of evidence in a baseline risk assessment is inappropriate, will result in an overly conservative estimate of risk, will dilute the interpretation of more scientifically defensible and site-specific measures, and will serve as an inappropriate basis for evaluating remediation options. The relative value of SQGs as one line of evidence in a weight-of-evidence approach is so low that it is negligible when compared to site-specific empirical data. Indeed, at the May 6, 1999 meeting, EPA recognized this and indicated that it was highly unlikely to use these SQGs to assess effects. In these circumstances, measurement endpoints based on comparisons to SQG should be omitted from the work plan. Furthermore, none of the SQGs qualify as applicable or relevant and appropriate requirements (ARARs); therefore, obligatory consideration is not warranted.

It should also be noted that the work plan also includes comparisons of surface water concentrations to ambient water quality criteria (AWQC) as a measurement endpoint for survival, growth and reproduction in fish, despite the inclusion of several site-specific studies on reproduction in fish. Although it is appropriate to recognize AWQC as ARARs, these criteria are generic and do not reflect site-specific risk. Therefore, studies providing direct evidence of site-specific risk to aquatic species should be given more weight in the risk analysis than comparison with the AWQC.
Parameters to be used in food chain modeling to estimate dose are unreasonable and unrealistic, or are not specified.

The work plan lacks sufficient detail describing the planned food chain modeling. The work plan should be revised to specify: (a) each receptor species to be modeled; (b) whether probabilistic, central tendency, high end, and/or maximum exposed individual estimates of dose will be estimated for each species; (c) the basis for each species' assumed diets and fractions ingested from the affected area; and (d) the fractions assumed bioavailable and absorbed. One of the few cases in which model parameters are specified is in the use of the maximum concentration or 95 percent upper confidence limit (UCL) on the mean to approximate chemical concentrations in environmental media. Use of the maximum detected concentration in an area or the 95% UCL does not reflect a realistic or representative exposure point concentration for receptors that utilize the entire areas. Rather, spatial averages (with 95 percent confidence intervals to quantify uncertainty) should be used, because they represent a more defensible description of actual exposure patterns.

Indeed, for all model parameters, reasonable and realistic assumptions should be employed (e.g., the assumed body weight should be the mean reported value and should be consistent within a single equation, the area use factor should be based on actual information on the receptor's home range sizes and the extent of available affected habitat). The work plan should also address how to avoid problems related to compounding conservatism in modeling (Cullen, 1994). Unless the changes suggested above are made, the model outputs will either be of low quality and/or overly conservative, resulting in model predictions that will not be reliable.

Furthermore, even if reasonable and realistic exposure assumptions are used in the modeling, the use of the model predictions to compare to literature-based RTVs in a TQ approach is still a conservative screening-level approach that is typically focused on individual organisms. Hence, if that approach predicts effects, further steps should be undertaken to evaluate the potential for impacts at the population or community level, as discussed in Section 2.7.
2.10 The majority of proposed field studies do not specify criteria used to select reference areas, the locations of candidate reference areas, or whether reference areas will even be evaluated. For those studies that do acknowledge that reference areas will be evaluated, the number of reference areas is not sufficient to characterize natural variability in the parameters under evaluation.

The proper definition of reference areas (and use of a sufficient number of reference areas) provides a critical line of evidence when establishing causality for any observed differences in measurements between the study site and the selected reference areas. Appropriate reference areas are critical to distinguishing actual effects of COPCs from effects of other factors, such as habitat extent and quality, predation, human development, and chemicals not related to the site. Inadequate definition of reference areas either will preclude any conclusions regarding potential adverse effects or will lead to erroneous or ambiguous conclusions concerning risk. In order to ensure that selected reference areas are appropriate, objective criteria for selecting references areas should be defined in the work plan a priori, and multiple candidate reference areas should be screened based on those criteria.

Although reference areas are vaguely mentioned on page 7-1 of the work plan ("areas upstream of the facility, outside of the drainage, and in Housatonic River tributaries will be used as reference areas for different aspects of the ERA"), with few exceptions (e.g., waterfowl), specific reference areas to be used for each measurement endpoint are not identified. All study designs should be revised to explicitly name the reference areas that will be evaluated. In the rare cases where the work plan currently specifies reference areas (waterfowl tissue residue, benthic invertebrate community structure survey and toxicity tests, and tree swallow reproduction survey), either the proposed reference area is inappropriate, or insufficient justification is provided to demonstrate that the reference area is appropriate. For example, Reach 1 is identified for the benthic invertebrate and swallow studies, but Reach 1 is unlikely to provide sufficient area or habitat to support viable populations that would be appropriate to compare to other sites. In cases where the work plan recognizes that reference areas will be evaluated, consideration of only a single reference area is generally proposed. Finally, given that no single reference area will exactly match the study area with respect to all of the key criteria, evaluation of multiple reference areas will minimize the influence of confounding factors in interpreting the study results. Therefore, the work plan should also be revised to note that more reference areas will be evaluated for each of the proposed studies. The number of reference areas specified for each study should be based on an evaluation of the habitat differences that occur in the Housatonic River study area.
2.11 Some of the studies, particularly the in situ toxicity tests, are inappropriate because they are not interpretable and will not add useful data to other proposed studies.

For example, there are no standardized protocols for conducting in situ tests and the relationship between such tests and standardized toxicity tests (e.g., 10-day sediment toxicity tests) is undetermined. Furthermore, as acknowledged in the work plan (App. A.4, p. 13): "[n]o standardized interlaboratory variance or power analyses have been conducted with these assays." In situ toxicity tests are in early developmental stages and their conduct would be of a research nature, especially given the planned conduct of laboratory toxicity testing and studies of indigenous benthic invertebrates. In addition to the inherent limitations of the in situ tests, the study design protocol is of such limited scope that the resultant data would not be interpretable given the stated objectives. The plan is to conduct such tests at only three sites in the Housatonic River, which is inadequate for determining stressor-response curves and no-effect levels. The work plan proposes to conduct an in situ evaluation of amphibian reproduction in vernal pools, despite the fact that a far more rigorous study is proposed to evaluate reproduction in leopard frogs collected from the Housatonic River and from a reference area under controlled laboratory conditions. Because the in situ frog study will likely be associated with far more uncertainty than the laboratory toxicity test, it is not expected to provide any additional information not already provided by the laboratory test. The work plan should be revised to omit studies that are not interpretable and/or do not provide additional useful data.

2.12 Although Section 7 of the work plan repeatedly states that an effort will be made to define no-effects levels and stressor-response curves as part of the individual studies, many of those study designs are not adequate to meet these objectives.

During the May 6, 1999 meeting, EPA clarified that many of the studies described in Section 7 are not intended to meet the identified objectives and that Section 7 is in error. Because no-effects levels and stressor-response relationships are critical to understanding causality and identifying site-specific effects thresholds, it would have been more useful to modify the study designs so that they could define no-effects levels and stressor-response relationships, rather than simply eliminate those objectives. For example, in some cases where collection of site-specific data is proposed in the work plan, additional sampling stations would help establish a concentration gradient necessary to support

---

1 Although Section 7 specifies that development of stressor-response curves and no-effects levels is a goal of this study, during the May 6, 1999 meeting EPA stated that this is not in fact a goal of this study.
these objectives (e.g., benthic macroinvertebrates, fish, mammals). All study designs presented in the work plan should be revised to: (a) support the development of no-effects levels and stressor-response curves; (b) demonstrate the feasibility of developing no-effects levels and stressor-response curves; and (c) explicitly describe the statistical methods for developing no-effects levels and stressor-response relationships.

2.13 The proposed exposure assessments do not address requirements for spatial data and spatially explicit decisions during risk management.

Many of the proposed studies do not adequately address the collection of spatial data. For example, there is inadequate spatial coverage specified in the benthic macroinvertebrate and sediment toxicity surveys, while the spatial coverage specified for tree swallows appears biased. All of the tree swallow nest box sites are located north of Woods Pond. Different nest box clusters are located close enough to one another so that foraging ranges of birds associated with different nest box clusters may overlap. There is also no discussion of the relationship between the exposure data that will be collected, the spatial aspects of the water/sediment quality, and bioaccumulation modeling. The work plan does not specify criteria for defining area use factors applied in wildlife exposure models to be conducted in support of TQs. Given all of these deficiencies, the ERA will not likely yield relevant information for making remedial decisions, and risk managers may be tempted to make decisions at very large scales across entire reaches of the river, unless the work plan is substantially revised. Specifically, the work plan's approach to exposure assessment should be modified to describe how the data will be treated for a spatial analysis of risk and to demonstrate that sufficient data will be collected to make such an evaluation. The work plan should describe the specific method that will be employed to organize data into “spatially relevant groups” (p. 7-24).

2.14 There is no analytical or interpretative framework for use of total PCB or Aroclor data vs. PCB congener-specific data.

The work plan indicates that EPA will obtain congener-specific PCB data on all biota sampled and will use those data to evaluate risks to such biota through a “TEF/TEQ” approach. Under that approach, certain PCB congeners are converted to toxic equivalents (TEQs) of dioxin through the use of toxicity equivalency factors (TEFs), and the risks are then evaluated on the basis of such TEQs. Although the work plan indicates that this TEF/TEQ approach has recently been accepted for use in
assessing risks to wildlife, GE is not aware that EPA has formally approved this approach for ERA.
and requests a reference that documents EPA’s acceptance.

The work plan also fails to provide details concerning how the TEF/TEQ approach will be applied and
does not indicate which species or toxicity endpoints will be evaluated using that approach. EPA does
not identify the specific PCB congeners to be analyzed in sediments, soil, surface water, and tissues.
Nor does it discuss the TEF scheme that will be used to estimate TEQs. The work plan does not
describe how the non-Ah receptor mediated effects will be evaluated or how the TEQ-based effects
thresholds will be developed.

In fact, TEF/TEQ approach is not a scientifically valid approach for evaluating risks associated with
PCB exposure for all species or for all toxicity endpoints. Therefore, this approach is not universally
applicable, and must be applied with considerable caution. For example, reproductive effects of PCBs
in birds and mammals appear to be largely Ah receptor dependent and may be predicted on a TEQ
basis. Conversely, results of field studies with fishes indicate that expression of PCB exposure as
TEQs does not improve correlations with adverse effects, compared to use of total PCB data. The
relative potency of individual congeners can vary considerably among fish species; consequently,
TEFs developed for any species may not be applicable to other species. Mac et al. (1993) investigated
the correlation of reproductive effects in lake trout with total PCBs, TEQs based on mammalian TEFs,
and TEQs based on lake trout TEFs. The authors found that the strongest correlation was with total
PCBs. Finally, the toxicology of PCBs in reptiles and amphibians has not been extensively
investigated and TEFs have not been developed for these taxa.

In general, there are three possible approaches for evaluating risk from PCBs: total or Aroclor-specific
PCBs, TEQs calculated using TEFs, and TEQs derived using a bioassay. All of the risk assessment
approaches for PCBs have both advantages and limitations. Using total or Aroclor-specific PCBs is
the most straightforward and least expensive approach. A limitation of this approach is that various
physical and biological factors can alter the composition of a PCB mixture once it is released into the
environment. This altered mixture might not be toxicologically equivalent to the original mixture.
Use of Aroclor-specific PCB data in a risk assessment assumes that the PCB mixture in the
environment has the same toxicological properties (e.g., toxic potency, bioavailability) as the
commercial PCB mixture used in the laboratory experiment from which the TRV was derived.
Both remaining risk assessment approaches for PCBs are congener-based and involve the expression of PCB concentrations as 2,3,7,8-TCDD TEQs. One approach converts PCB data to TCDD TEQs using TEFs, and the other approach uses a bioassay. Neither approach is ideal. The major limitations associated with a TEF-based approach are related to assumptions inherent in the approach, which include the additivity of toxicity for congeners in mixtures, no variation in sensitivity across taxa, parallel dose-response curves, and the toxicity of PCBs being solely attributable to the dioxin-like congeners. There are published exceptions to all of these assumptions, which can limit the validity of this approach (Safe, 1990, 1994, 1997, 1998). Bioassay-based approaches to deriving 2,3,7,8-TCDD TEQs that address the behavior of congeners in site-specific mixtures and species-specific sensitivities can be developed. A bioassay approach, however, addresses the effects of dioxin-like substances only and does not distinguish between dioxin-like PCBs and other Ah receptor active compounds, such as certain dioxins, furans, and polycyclic aromatic hydrocarbons (PAHs). Therefore, chemical analysis of these latter compounds may be required to delineate their relative contribution to the total TEQ.

Thus, although the TEQ approach provides a convenient means to evaluate ecological risk on a congener-specific basis, it must be used with considerable caution. The results of wildlife and laboratory animal studies suggest that the predictive value of TEQs for PCBs is species-, response-, and mixture-specific (Safe, 1994). Therefore, these parameters must be considered when selecting a unit of expression for assessment of PCB exposure and toxicity (i.e., total PCBs or TEQs) in an ERA. A TEQ approach should only be considered if the response being predicted is mediated through the Ah receptor and the relative sensitivity of the receptor (or a closely related species) to dioxin-like compounds is known.

In short, the TEF/TEQ approach should not be used for fish, and the work plan should be revised to provide more detail on the use of that approach for other receptors, including specific congeners to be evaluated, the TEF scheme to be used, which receptors/endpoints will be evaluated using this approach, and the rationale for using that approach.

2.15 Many of the non-standardized test protocols and field work appear to be research oriented and not directly applicable to the estimation of ecological risks and risk management.

Several of the proposed studies (e.g., the long term and in situ sediment toxicity tests, fish egg injection study, and endocrine disrupter study) appear to be based on methods that are currently in development, lack field validation, and/or lack a causal link to adverse effects at even the individual
level, let alone the population or community level. Watershed-scale ERAs such as that proposed for the Housatonic River may appear to present an ideal opportunity to address research needs for the evaluation, development and advancement of potential assessment methodologies. However, the purpose of the investigation and ERA is to gather the data necessary to develop an accurate understanding of the environmental conditions and the risks posed to the ecological community by chemical exposure for risk management decision making purposes. The regulatory process should not be used to address research and development needs identified within EPA or academic circles.

2.16 The extent of the sediment sampling and chemical analysis programs proposed appear to be out of proportion with level of effort for most of the ecological risk studies.

The proposed sediment sampling program requires the collection and chemical/physical analysis of more than 5,000 sediment samples from the Housatonic River. In contrast, some of the ecological studies (e.g., crayfish bioaccumulation, invertebrate and frog toxicity testing) on which the ERA will be based are limited in that they address only a limited portion of a single reach of the river, lack a sufficient number of sampling locations within a given reach to address the variability in habitat and exposure conditions, and lack sufficient sample replication at each location for physical, chemical and biological analyses. This imbalance suggests a focus on remediation prior to completion of the ERA. Furthermore, the inadequate design of the ERA may lead to uninterpretable data, which in turn may lead to a dependence on the conservative screening criteria methodologies. A more appropriate balance needs to be established for the level of effort expended for determination of chemical extent versus assessment of ecological risk.

2.17 The protocols presented in the appendices lack adequate detail and/or are poorly designed.

Numerous flaws are common to the vast majority of protocols presented in Appendix A and B. First, the protocols are either ambiguous or do not discuss how results will be used and interpreted (e.g., will the population surveys be used for descriptive purposes only or will they be compared to results from reference areas in order to evaluate risks). The sample sizes, timing and frequency of sampling events are either not specified or not justified (e.g., benthic macroinvertebrates, fish). Most appendices lack information on treatment of data, such as quality assurance/quality control (QA/QC), data reduction, presentation, statistical treatment, and application. The lack of detail provided prevents the complete evaluation of the feasibility, quality and scientific defensibility of many of the proposed studies. Consequently, there is the potential for substantial subjectivity in the interpretation of data generated.
Furthermore, these descriptive gaps in the protocols may cause studies to be improperly executed, and some critical data may not be collected. Unless the study designs are significantly improved (based on the many specific comments provided in Section 3.0 below), the proposed studies will result in data that are ambiguous or irrelevant to the ERA, which in turn will lead to subjective interpretation of the results. It is worth noting that the notable exceptions to many of these gaps are Appendices A.9 and A.12, which are of substantially higher quality than virtually all other appendices. All other appendices should be rewritten, closely following the example of the level of detail and content provided by Appendices A.9 and A.12. Numerous specific examples of flaws in the appendices are presented in the following section.

3.0 SPECIFIC COMMENTS

This section presents detailed comments on those portions of EPA's work plan that pertain directly to investigations associated with the ERA. Consequently, the following comments do not address the sampling of algae, phytoplankton, zooplankton, macrophytes, PMI, BMI, bioturbation, and porewater, all of which are being conducted for use in the bioaccumulation model (which is part of the overall fate and transport model). Comments are organized by receptor group, and pertain to both volumes of the work plan (i.e., the overview text provided in Volume I and the specific study protocols provided in Volume II).

3.1 Rare Plants and Natural Communities

Overall, the plan for evaluating rare plants is adequate and should succeed in locating any rare plants or rare plant communities in the study area. The plan follows a logical progression and includes preliminary work to determine what rare plants/communities may exist in the areas and identify locations with the greatest potential for supporting rare plants/communities. However, more specific information should be provided in the plan concerning actual field methodologies and personnel qualifications.

While the use of landscape analysis is a good tool to identify potential habitat for rare species or communities, it is unclear whether, once identified, these locations will be the only ones surveyed. Although the majority of the time should be focused on these high-potential locations, the entire area should be evaluated from the ground. The lower-potential areas should be driven through to evaluate whether there are any small pockets of habitat not observable from the aerial photos, topographic
maps, or other available information. If present, they could be evaluated. If not present, then these areas could be checked off the list and eliminated from further consideration. Otherwise, it is possible that some areas could be missed.

The study design states that "attempts will be made to perform surveys during periods when individual species can be positively identified ..." This is crucial for the correct identification of some species. There is no mention, however, of the frequency of the surveys or how much time is going to be devoted to conducting the rare plant/rare community surveys throughout the growing season. It is unlikely that all the potential rare species are going to bloom at the same time during the summer. Therefore, multiple visits will be required throughout the growing season. Depending on the species present, the areas with highest potential for rare plants should be visited biweekly throughout the entire growing season.

The field methodology to be used to search for rare species at the high-potential locations has not been discussed. A systematic approach using a grid or transect lines is necessary to ensure the entire area is surveyed. No mention is made of the qualifications of the field botanists who will conduct the work. It is very important that the personnel conducting the field survey are qualified botanists, familiar with rare species and the communities that they are looking for. Otherwise, it is likely that some species or communities will be missed or overlooked. Also, there is no indication of the number of field personnel or the qualifications of these individuals. The role of assistants with respect to the actual searches for rare species should be clearly identified.

It is strongly suggested that the locations of all rare plants and rare plant communities that are found be surveyed using global positioning system (GPS) equipment. Modern equipment and resources make this an efficient task and allows for very accurate placement on maps and relocation of populations in the future, if necessary. The rare plants and communities program should be revised based on the comments above.

3.2 Dragonflies

This survey is intended to characterize the distribution of dragonflies and the occurrence of rare species along the Housatonic River, and therefore will be of limited use in the ERA. The objectives of this study will be accomplished through a literature review and personal contacts to determine
historical distribution of dragonflies in the Housatonic River drainage system. In addition, exuviae (on riverbanks) and adults will be collected during five sampling periods from May through September.

In describing the dragonfly survey, the work plan (p. 5-68) states that “[t]his information will be used in the ecological risk assessment to assess community structure.” However, the description of the survey in Appendix A.3 indicates that only the occurrence of dragonfly species, not their relative abundances, will be determined. If the study plan is accurate, an assessment of community structure will not be possible using the data collected during this survey.

Methods for geographic positioning (e.g., GPS) for mapping of dragonfly surveys and captures are not specified (Appendix A.3). Because the primary purpose of the survey is to document the presence and locations of dragonflies collected, precise documentation of captures is important. The dragonfly program should be revised based on the comments above.

3.3 Crayfish Bioaccumulation

Concentrations of PCBs in crayfish will be determined to support the exposure assessment. However, there are several major limitations in the study design that are related to sampling time and number of stations. Each major limitation is described below.

The stage of gonad development in crustaceans can be a major factor determining the concentrations of organochlorines in tissues (Phillips, 1980). Because the characteristics of PCB uptake by crayfish can vary over the course of an annual cycle, it is essential that the time of sampling be selected carefully to ensure that the results will provide the appropriate kind of information for the exposure assessment model. The planned sampling period of May might be too early in the season to represent the peak periods of potential PCB uptake and therefore, may not be representative of the major periods of bioaccumulation.

Given the potential temporal variability in PCB uptake by crayfish, the use of a single sampling period for providing input to the exposure assessment should be justified. The collection of bioaccumulation data during several periods of the annual cycle would likely improve the predictive accuracy of the model. Only five stations will be sampled for the evaluation of crayfish bioaccumulation in downstream sections of the Housatonic River. To evaluate potential PCB uptake under the wide range of environmental conditions and sediment PCB concentrations found in the Housatonic River, more
habitats throughout the river should be sampled (see Section 2.10). In addition, a sufficient number of stations within each kind of habitat should be sampled to ensure that a wide range of sediment PCB concentrations is evaluated. The crayfish bioaccumulation program should be revised based on the comments above.

3.4 Benthic Macroinvertebrate Community Evaluations

The use of benthic macroinvertebrate communities to assess potential ecological risk is appropriate because the focus is on effects at the population and community levels, as recommended by EPA (1998a). In addition, evaluations of benthic macroinvertebrate communities have been used as a critical component of the Triad approach for characterizing sediment toxicity in numerous studies throughout the United States (e.g., Chapman et al., 1987, 1991, 1996; Canfield et al., 1996; Carr et al., 1996). This approach allows sediment toxicity to be evaluated using multiple lines of evidence, as recommended by EPA (EPA 1997a,b; 1998a, 63 FR 26889).

Despite the potential merits of evaluating benthic macroinvertebrate communities in the Housatonic River, the study design is flawed for numerous reasons and, therefore, will not provide an adequate assessment of benthic communities on the river. The major flaws in the study design are related to sampling time, number of stations, reference area characterization, incomplete Rapid Bioassessment Protocol (RBP) assessment, sediment sampling methods, RBP interpretation, and analytical methods.

Because the characteristics of benthic macroinvertebrate communities can vary substantially over the course of an annual cycle (Plafkin et al., 1989; Rosenberg and Resh, 1993; EPA, 1996), it is essential that the time of sampling be selected carefully to ensure that the results will satisfy the study objectives. This is particularly critical when sampling will only occur during a single period of the annual cycle. Examples of study objectives that influence sampling time are evaluation of “stable” communities and evaluation of colonization potential. If the study objective is to evaluate the “stable” community that persists in an area over time, sampling should occur during a period after seasonal colonization has occurred and selective pressures such as predation and physical/chemical environmental stressors have had their major effects on the communities. The surviving organisms can thereby be considered representative of the persistent or “stable” community in each area, free from the variability introduced by seasonal colonization and recruitment. If the study objective is to evaluate the colonization potential of an area, sampling should occur after the major period of colonization has occurred and organisms have had enough time to grow to a size sufficient for capture
by the target sieve mesh size. The results of this kind of sampling can be used to estimate the colonization potential of an area, even if most of the newly recruited organisms will not survive in the longer term. The evaluation of colonization potential can be used to estimate the value of an area for providing a forage base for fishes or larger macroinvertebrates during critical feeding periods.

According to EPA (1996), the optimal biological sampling period should be consistent with recruitment cycles of the benthic organisms, from reproduction to emergence and migration, so that the maximum amount of information can be derived from the data. EPA (1996) considers the optimal time of sampling as the period during which the target assemblages have stabilized after larval recruitment and subsequent to mortality. This is the period when use of the available microhabitats is maximized.

Plafkin et al. (1989) have also noted that the optimal biological sampling period should correspond to the recruitment cycles of the benthic organisms. The authors state that community information is maximized when most organisms are within a size range (e.g., late instars) retained during standard sieving and sorting, and can be identified with confidence. They conclude that the optimal sampling period is when the habitats are utilized most heavily by later instars and the food resource has stabilized to support a balanced indigenous community. The sampling should be conducted in late August or early September at a time when the communities are relatively stable and to allow synoptic sampling with the macroinvertebrate bioaccumulation analysis.

The work plan proposes use of artificial substrates to evaluate benthic macroinvertebrate communities. Due to several limitations, these samplers are unlikely to obtain a representative sample of the benthic macroinvertebrate community. First, these substrates are considered passive samplers that rely on the behavior of organisms to bring them in contact with the samplers so that they can be collected. This is in contrast to active samplers, such as kick nets and grab samplers, that capture organisms directly and are largely independent of the behavior of the organisms. Because artificial substrates do not realistically simulate the physical and chemical properties of natural substrates and they rely on the behavior of organisms for collection, they can result in highly selective and biased samplers. For example, the artificial samplers are selective for only certain kinds of benthic macroinvertebrates (APHA, 1989). Second, the effectiveness of the samplers can be strongly influenced by physical factors such as water velocity and installation depth (APHA, 1989; EPA, 1990). Third, artificial samplers provide no evaluation of the biota naturally present at a location, the condition of the natural substrate, or the potential effect of sediment contaminants (EPA, 1990). Fourth, because the samplers
only collect the organisms present during the six-week sampling period, they do not provide an integrative measure of potential long-term biological effects, as does the direct sampling of resident communities (EPA, 1990).

Given the numerous limitations of the artificial substrate samples identified above, they should not be used to provide an assessment of the potential effects of PCBs on resident benthic macroinvertebrate communities in the Housatonic River. The artificial substrates are also redundant because of the planned study of indigenous riffle communities (i.e., kick net samples). The kick net samples will also not provide meaningful information on the potential effects of PCBs because of the coarse substrates (e.g., gravel) and resultant low exposure to fine-grained sediments that may contain chemicals. Thus, there is a low potential for PCBs to adversely affect macroinvertebrate communities living in non-depositional habitats (i.e., riffle communities). Therefore, both the artificial substrate samples and studies of indigenous riffle communities should be deleted from the study design. Based on conversations with EPA, it is our understanding that EPA has agreed to drop the deployment of artificial substrates and the collection of riffle community samples. GE supports that decision.

The work plan states that, for each of the three kinds of major benthic habitats that will be sampled, only seven stations will be evaluated. This number is inadequate to characterize the study area, given the size of the area and the diversity of habitats within the area. Additional sampling stations located in the Housatonic River should probably be sampled in soft sediments, which are expected to have the highest concentrations of PCBs. The characteristics of benthic communities are strongly influenced by numerous physical/chemical habitat features [e.g., depth, flow, sediment grain-size distribution, and sediment total organic carbon (TOC) content] and biological features (e.g., proximity of macrophyte beds and predator abundance) (Plafkin et al., 1989; Rosenberg and Resh, 1993; EPA, 1996). Before the appropriate number of sampling locations can be determined, the various kinds of benthic microhabitats present in the study area should be determined and a sufficient number of stations should be used to ensure that all major kinds of microhabitats are sampled. In addition, sediment concentrations of PCBs within each major kind of microhabitat should be determined or estimated, and a sufficient number of stations should be used to ensure that a gradient of PCB concentrations is sampled within each microhabitat.

If a sufficient number of stations is used to account for habitat variability and variable sediment concentrations of PCBs, the probability of accurately estimating potential risks due to PCB exposure will be increased. If an insufficient number of stations is used, effects found for the benthic
communities may erroneously be attributed to PCB exposure when, in fact, they were due to differences in habitat characteristics.

The study design specifies that a single upstream reference area will be sampled for each of the three kinds of benthic habitats that will be evaluated. Although the use of reference areas to discriminate potential ecological risks in the study area is an appropriate technique (Plafkin et al., 1989; Rosenberg and Resh, 1993; EPA, 1996), the use of a single reference area for each kind of benthic habitat is wholly inadequate, given the range of microhabitats present in the study area (see Section 2.10). Based on conversations with EPA, it is our understanding that EPA may add one more reference station; however, this addition is still insufficient to fully characterize the full range of reference conditions. Because the reference areas will be used to discriminate potential effects due to PCB exposure, it is essential that they encompass the range of physical, chemical, and biological characteristics found throughout the study area. Otherwise, habitat-related effects on benthic communities may erroneously be attributed to PCB exposure. A comprehensive reference area evaluation should, therefore, be conducted in conjunction with the recommended habitat characterization of the study area. In this manner, a sufficient number of reference stations can be selected to account for the range of microhabitats found in the study area.

The RBP protocols for benthic macroinvertebrates were developed by EPA to provide an assessment of biological impairment in wadable freshwater streams (Plafkin 1989; Barbour et al. 1992), although EPA considers the approach also applicable to large freshwater rivers. A key component of a complete assessment is inclusion of samples of coarse particulate organic matter (CPOM), which primarily includes the plant debris (e.g., leaves, needles, twigs, bark) that accumulate in depositional areas. The CPOM samples typically include organisms that belong to the shredder functional feeding group, which may be underrepresented in the cobble substrate sampled by kick nets. The data from the CPOM samples are used to calculate the ratio of shredders to total individuals, which is one of the eight metrics calculated using the RBP protocols. Given the importance of CPOM samples to the RBP methods, it is essential that those samples be collected in conjunction with the kick-net samples planned for the study area. Otherwise, an important component of the benthic communities will not be sampled or included in the risk evaluation.

Although the use of a grab sampler to sample benthic communities in soft-bottom sediments is appropriate (APHA, 1989; EPA, 1990; Rosenberg and Resh, 1993), the failure to specify a minimum acceptable sampling depth can lead to an undersampling of the communities and erroneous
conclusions regarding potential ecological risk. Because benthic macroinvertebrates can inhabit sediment depths of 10–15 cm (e.g., ASTM, 1995; Charbonneau et al., 1997; Charbonneau and Hare, 1998), it is essential that the majority of this biologically active zone be sampled. Using this methodology, more organisms can be collected for use in the ERA.

Although grab samplers are generally considered effective tools for sampling benthic communities in soft-bottom habitats, they frequently sample different sediment depths among stations, as well as among the individual casts at each station (Nalepa et al., 1998). Variables that limit sediment penetration depth include the grain-size distribution of the sediment, as well as interfering objects such as stones, shells, and twigs. It is therefore essential that sediment penetration depth be standardized and monitored, and that casts that fail to meet the minimum specified penetration depth be rejected.

Typically, the ponar grab does not allow access to the surface of a sediment sample so that penetration depth can be determined accurately. However, the ponar grab can be modified to allow access by placing doors on its upper face. Alternatively, an Ekman grab could be used to collect samples, because that kind of sampler has doors on its upper face by design (APHA, 1989). Another method of standardizing penetration depth is to place the grab sampler on a pole that can be pushed a known distance into the sediment. This technique prevents complete reliance on the weight of the sampler to penetrate the sediment, which usually varies in relation to sediment grain-size distribution and the presence or absence of interfering objects. Another alternative is to use a weighted box core sampler with a lid, and to sample to a depth well below the target depth.

The metrics used in the RBP spell out protocols that were designed to evaluate potential biological impairment in general (Plafkin et al., 1989). The cause of any impairment found at a station could be due to numerous factors such as flow, substrate characteristics, and organic enrichment. The study design needs to describe how the various metrics will be used to distinguish and evaluate the effects of toxic chemicals on benthic communities.

Although the work plan provides detailed methods for the data analysis and interpretation of the riffle samples (i.e., the RBP metrics and protocols), it does not provide specifics on how the soft-bottom community data will be analyzed and interpreted. In addition, no details of the Triad data interpretation procedures are specified. This general lack of consideration for analytical or statistical procedures makes the validity of the study design questionable, because there are no criteria by which the adequacy of the numbers of stations and replicate samples can be judged. As an example, the
number of replicate samples used for statistical comparisons should be based on some form of statistical analysis to ensure that the data collected in the field can support the planned analyses (e.g., Green, 1979). There are also internal inconsistencies in the work plan concerning numbers of samples that will be collected at each location. The numbers of discrete samples at each benthic study location are specified in different places as three (p.7-37) and as five (p. 5-70).

In addition, because the Triad evaluation typically requires that all data be collected synoptically, it is critical that all of the Triad analyses be specified prior to data collection. Otherwise, resampling may be required to collect information on missing variables. This resampling would introduce uncertainty into the relationships among the Triad indicators due to sampling variables such as station location and seasonality. Although GE supports the use of the Triad approach in general, the lack of specific details in the planned approach results in considerable uncertainty regarding the appropriateness of the planned data analyses. Moreover, because of the limited number of sampling stations (discussed above), meeting the stated objectives (i.e., to develop stressor-response curves and no-effect levels) may be severely compromised. The Triad data interpretation is further limited by the specification of only four stations (one reference and three test sites) for the sediment toxicity tests. Thus, the full Triad data set for the Housatonic River would only include three stations. Development of stressor response curves and no-effects levels with such a limited data set would be highly uncertain at best, and most likely impossible.

The macroinvertebrate community analysis program should be revised substantially based on the comments above.

3.5 Sediment Toxicity Testing

The use of sediment toxicity testing to evaluate potential risks of toxic chemicals in sediment is an appropriate methodology, as well as a component of the Triad approach to sediment assessment (Long and Chapman, 1985; Chapman et al., 1987; Chapman, 1990; Chapman et al., 1997). The sediment Triad approach has been used to characterize sediment toxicity in numerous studies throughout the United States (e.g., Chapman et al., 1987, 1991, 1996; Canfield et al., 1996; Carr et al., 1996). This approach allows sediment toxicity to be evaluated using multiple lines of evidence, as recommended by EPA (EPA, 1997a,b; 1998a; 63 FR 26 889). The test species selected for the evaluation of sediment toxicity in the Housatonic River (i.e., Hyalella azteca and Chironomus tentans) have been
shown to be sensitive indicators of sediment toxicity in standardized 10-day exposure period tests (Burton and Scot, 1992; Ingersoll et al., 1995; EPA, 1994; ASTM, 1995).

Despite the merits of evaluating sediment toxicity in the Housatonic River, the methods of toxicity testing identified in the work plan are considered incapable of providing an adequate assessment of toxicity with any degree of certainty. The major shortcomings of the study design are related to the number of stations, reference area characterization, test methods, and redundant studies.

Only four stations will be sampled for sediment toxicity testing. This small number of stations will not adequately characterize the study area, given the size of the area and the diversity of sedimentary environments within the area. Because PCB concentrations vary substantially throughout the study area, more stations are needed to assess sediment toxicity adequately. In addition, factors that modify PCB toxicity, primarily sediment TOC content (e.g., DiToro et al., 1991), vary throughout the study area. Four stations are also inadequate to identify any kind of meaningful stressor-response curve between PCB concentrations and toxicity endpoints, or to determine no-effect levels.

As is the case for the benthic studies, the study design for sediment toxicity testing specifies that only a single reference station will be used. This is inappropriate for several reasons. First, although use of a reference station to determine the significance of any sediment toxicity in the study area is an appropriate technique (EPA, 1994, 1997a), the use of a single station will not allow the response variability of the test organisms to be fully evaluated for a range of reference conditions (see Section 2.10). Based on conversations with EPA, it is our understanding that EPA may add one more reference station. Although the addition of one reference station will improve the study design, the overall characterization of reference conditions will still be inadequate. The evaluation of reference area variability is critical because it directly affects the determination of significant toxicity in the study area. Response variability is likely to be associated with the different physical/chemical characteristics of the sediment, and therefore can influence the responses of the test organisms. For example, most test species are sensitive to some degree to sediment grain size distribution and TOC content (EPA, 1994; ASTM, 1995). This sensitivity arises in part from the fact that the test organisms usually are not native to the reference sediment and are therefore placed in a foreign environment without any kind of acclimation. Even though the reference sediment has low concentrations of toxic chemicals, its physical/chemical characteristics alone can be stressful to the test organisms, and thereby increase the variability of the response to the sediment. It is therefore critical that more than one or two reference stations be evaluated.
Second, by evaluating multiple reference stations, the range of non-chemical factors that affect toxicity responses can be better characterized. This effect is critical when conditions in the study area vary widely, as they do in the Housatonic River. At a minimum, multiple reference stations should be used to bracket the ranges of sediment grain size distribution and TOC content found in the study area.

None of the proposed test methods are standardized or are commonly used in sediment assessments (see Section 2.15). Although draft standardized protocols are currently under review, they have not been widely tested for consistent performance and generation of meaningful results. Instead, the protocols are largely experimental in nature and therefore potentially subject to unknown biases and experimental artifacts. These potential problems could compromise the validity of the test results and lead to erroneous conclusions regarding the potential role of PCBs in sediment toxicity. The lack of standardized protocols and widespread use of the methods specified in the study design are in direct contrast to the protocols for the 10-day tests based on *Hyalella azteca* and *Chironomus tentans* (e.g., EPA, 1994; ASTM, 1995).

In contrast to the standard 10-day tests, the long-term toxicity tests using *Hyalella azteca* and *Chironomus tentans* do not meet EPA’s (1994) selection criteria for toxicity tests with respect to known interlaboratory variability and field validation. It is recognized that interlaboratory variability has been evaluated for these tests, but the results have not been published and were not available for this review. Therefore, the test results could be biased to an unknown degree by the manner in which the laboratory implements the protocols and the ecological relevance of the test results would be unknown. In selecting the toxicity tests recommended for assessing the toxicity of freshwater sediments, EPA (1994) identified 12 criteria that are considered essential for an appropriate test. Although the two test species meet all of the criteria specific to selection of the test species, the long-term protocols do not meet the following major criteria specific to the test methods:

- The test should have a database for interlaboratory comparisons of procedures (e.g., round-robin studies)
- The test should be confirmed with responses in natural populations of benthic organisms.

Although interlaboratory comparisons for the long-term chironomid test are apparently in progress (EPA, 1997c), the draft protocol for the long-term amphipod test states in Section 14.1.1 that “only a limited number of laboratories have used the method.” and “more definitive methods may be
described in future versions of this manual after additional laboratories have successfully used the method."

With respect to field validation, Section 14.4.7 of the amphipod protocol and Section 15.4.7 of the chironomid protocol state that "research is also needed evaluating the ability of these laboratory endpoints to estimate responses of benthic organisms exposed in the field to contaminated sediments." This statement clearly indicates that the protocols are not ready for routine application because the ecological relevance of the test results is not adequately known.

An additional factor that adds uncertainty to the results of the chronic tests is the ability to hold the test organisms under laboratory conditions for 40 to 60 days without creating stress due to factors other than chemical toxicity. If the organisms were stressed due to the artificial holding conditions, the test results could be biased from the direct effects of the stress, or because the stress made the organisms unusually sensitive to chemical toxicity. Subtle endpoints such as growth and reproductive effects would likely be most susceptible to stress-induced biases.

A particularly disturbing specification of the long-term protocols for both test species is that the level of dissolved oxygen (DO) in the overlying test water is allowed to decline to 2.5 mg/L before aeration is required (EPA, 1997c). This value is approximately 29 percent of saturation at the test temperature of 23°C, which is well below the minimum acceptable level of 40 percent of saturation specified for the 10-day tests (EPA, 1994; ASTM, 1995). Because stressful conditions due to low concentrations of DO can interfere with the assessment of chemical toxicity, the subjection of the test organisms to long periods of low DO concentrations could confound the interpretation of the test results.

In Section 14.4.7 of the amphipod protocol and Section 15.4.7 of the chironomid protocol, it is stated that additional studies are needed to further evaluate "the influence of grain size and ammonia on long-term exposures." This statement clearly indicates that the protocols are not ready for routine application, because they may be subjected to experimental artifacts.

Finally, there is no substantial database with which to compare the results of the chronic toxicity tests. This is in contrast to the large national database for the 10-day amphipod and chironomid tests. Because of this limitation, the validity and meaning of the results of the proposed chronic tests cannot be verified by comparing them to results from other studies.
3.6  *In Situ Toxicity, Bioaccumulation, and Stressor Identification Studies*

The work plan calls for several kinds of *in situ* exposures that are intended to evaluate toxicity, bioaccumulation, and stressor categories in the Housatonic River. When viewed as a group, these proposed studies suffer from the lack of standardized procedures, field validation, and existence of a comparable database. Many of these procedures are in developmental stages and are of a research nature. They are therefore inappropriate for a definitive ecological risk assessment that may form the basis for remedial decisions. As a group, these test procedures do not have valid controls that could be used to separate stresses related to the artificial deployment from intrinsic stresses that may occur in the Housatonic River under natural exposure conditions. Finally, because these tests are conducted in the field under uncertain exposure conditions, they are subject to significant environmental variability that cannot be controlled or even monitored.

The *in situ* toxicity tests will not provide meaningful information concerning the toxicity of chemical stressors to indigenous organisms. The proposed tests are redundant with the benthic macroinvertebrate community evaluations and the laboratory sediment toxicity tests. In addition, the *in situ* protocols are inferior because they are not standardized and have not been widely used. If the objective of the *in situ* toxicity studies is to evaluate sediment toxicity, the use of 10-day laboratory toxicity tests and evaluations of benthic macroinvertebrate communities is more appropriate because those methods have been standardized and widely used throughout the United States (e.g., Chapman et al., 1987, 1991, 1996; Canfield et al., 1996; Carr et al., 1996). The endpoints (survival) for two of the proposed *in situ* tests (*Daphnia* and *Hyallela*) are also less sensitive than the proposed or alternative laboratory toxicity tests (e.g., growth). The exposure times for the *in situ* toxicity tests are also considerably less than the proposed laboratory tests. Therefore, the *in situ* toxicity tests would be assessing completely different exposure and response conditions when compared to the laboratory tests. In all likelihood, the two kinds of tests would have very different results and the work plan presents no interpretive framework for dealing with such results. Therefore, the *in situ* toxicity tests should be deleted from the study design.

The first proposed component of the *in situ* studies is the evaluation of certain test species for bioaccumulation of PCBs in tissues. This *in situ* bioaccumulation study is redundant with the tissue analyses that will be conducted on indigenous benthic macroinvertebrates. The work plan includes substantial sampling of indigenous biota representing various biological groups. If properly designed, such studies will assesses bioaccumulation under realistic exposure conditions. Its application to benthic macroinvertebrates is particularly useful because most species are relatively sedimentary,
Unlike fishes, and therefore spend all or most of their life cycle in the assessment areas. The use of artificial in situ exposure conditions can introduce many uncertainties and biases into the evaluation and, therefore, would provide redundant information of lower quality with respect to overall data interpretability. Hence, the in situ bioaccumulation study should not be conducted.

The second proposed component of the in situ studies listed in the work plan is the deployment of Stressor Identification Evaluation (SIE) chambers at the toxicity test sites. The SIE test methods are not standardized, use very short exposure periods (two days), use insensitive endpoints, and have a high potential for yielding equivocal results. These proposed tests are clearly of a research nature and should not be conducted as part of an ERA. Based on conversations with EPA, it is our understanding that EPA has agreed to drop the SIE studies. GE supports that decision.

In this connection, the work plan notes that laboratory toxicity identification evaluation (TIE) tests will also be conducted with sediments from each site. These tests also suffer from major limitations and do not have final EPA guidelines available. The TIE tests are designed to separate the potential effects of four stressor categories: PAH photoinduced toxicity, ammonia, metals, and nonpolar organics. They are not designed to separate potential toxicity of individual metals or individual nonpolar organic substances. The tests are also based on exposure of planktonic organisms (Ceriodaphnia or Daphnia) to sediment pore water. Thus, the exposure conditions are highly unrealistic and the results are not necessarily applicable to indigenous benthic fauna. For these reasons, the TIE tests should not be conducted.

The final component of the proposed in situ studies listed in the work plan is the deployment of semi-permeable membrane devices (SPMDs). Although these devices have been used as a surrogate for tissue bioaccumulation analyses, the results are not directly comparable to bioaccumulation in any specific indigenous organism. They may be used to monitor temporal trends in water quality when individual species cannot be collected at all times or sites. However, SPMDs are redundant and have uncertain applicability when indigenous organisms can be collected for bioaccumulation studies. The work plan also indicates that SPMDs will be used to evaluate possible sources of PCBs from the groundwater pathway. However, the work plan indicates that SPMDs will be deployed at only three unspecified sites on the Housatonic River. With such limited spatial sampling, the potential for any meaningful information concerning PCB sources is extremely low. Thus, SPMD deployment is unnecessary and should be deleted from the work plan. Based on conversations with EPA, it is our
understanding that EPA has agreed to drop the deployment of SPMDs from the work plan. GE supports that decision.

3.7 Freshwater Mussels

The work plan describes two studies involving freshwater mussels, including a survey of mussel presence/absence (Appendix A.5) and an in situ bioaccumulation/growth study (Appendix A.6). There are significant inconsistencies between these two protocols. The work plan presented in Appendix A.6 indicates that the freshwater mussel survey was conducted in 1998 and found "almost a complete absence of mussels" (App A.6, p. 1). However, the work plan for the mussel survey presented in Appendix A.5 is dated February, 1999 and makes no mention of such a survey. It appears that either the protocol for the mussel survey was completed after the survey was conducted or that the authors of the protocol were unaware of the previous survey. In either case, this inconsistency does not inspire confidence in the quality of the proposed study (see Section 2.2). Based on discussions during the May 6, 1999 meeting, it is our understanding that a mussel survey was indeed conducted in 1998 and that a second survey is being conducted and the design of that survey has been modified based on the results of the first survey. Neither results of the first study nor the revised study have been released to GE. There is therefore no opportunity for review or comment on the design of either study. It was also indicated at the May 6, 1999 meeting that the in situ mussel study was designed to determine the cause of reduced mussel populations in the Housatonic River. No information was provided as to how the in situ study would be related to the field surveys. Moreover, it is unclear that the endpoints being measured in the bioaccumulation study could be used to extrapolate to mussel populations in the watershed. Specific concerns related to each of these studies are described in the following sections.

3.7.1 Freshwater Mussel Inventory

As discussed in Section 2.1, it appears that the mussel survey has already been conducted; therefore, comments and/or concerns identified by GE (and others) cannot be addressed unless the entire study is either redone or omitted from the ERA. Deficiencies of this study are described below and include poorly specified and inadequate reference sites, and a flawed study design (See Section 2.1).

First, the survey design is inadequate to assure a rigorous evaluation of mussel populations in the Housatonic River. Specifically, the inadequate use of reference sites means that there is no basis for comparison with other similar systems in the Northeast. This is particularly important for freshwater mussels that have been in decline throughout New England since the late 1800s. Only through the use
of carefully chosen reference sites can site-specific differences in mussel populations be distinguished from this general decline.

Second, despite the fact that this survey has apparently been completed, the survey design does not specify the locations of target and reference areas that were surveyed. The study areas that are referred to are the Newell Street to the Woods Pond study area, and upstream and downstream of the study area. A single reference site is proposed in the study design, although its location is not identified. A single reference site is inadequate to define the range of variability in water temperature, food sources, disturbances, and sedimentation that will be evaluated along the river.

Third, the selection of study areas cannot be supported by the study design. Habitat characterization was proposed only where mussels were found, rather than basing the selection of study areas on habitat suitability. Without a detailed description of the methodology to be used to locate mussels (e.g., based on transects or habitat), it is not possible to judge the adequacy of the survey; an inappropriate search strategy could lead to the underestimation of the population. In addition, the survey, as described in Appendix A.5, was to be limited to a depth of three feet using a viewing scope, rather than to an appropriate depth based on habitat using a snorkel or SCUBA survey. The thoroughness of the inventory will, therefore, be of limited value.

Finally, it is important to consider that mussel populations can be influenced by a number of factors. Although the authors present a limited list of such factors (App. A.6, p. 1), there is no provision in the study design for evaluating the factors that might contribute to the low numbers of mussels found in the 1998 survey (results described in on App. A.6, p. 1). For example, despite their specific inclusion in the objectives, there is no provision in this work plan for the evaluation of species that would act as the host required for the larval stage of mussel development or for the identification of wildlife species that might prey on mussels.

As previously indicated, the distribution of mussels in Northeast river systems has been in decline over the past century and the Housatonic River should be considered within this context. Given the problems described above, data collected from this study will be inadequate and unusable in the ERA.
3.7.2 Freshwater Mussel Bioaccumulation and Growth Study

Based on discussions during the May 6, 1999 meeting, it is our understanding that there have been significant changes to the protocol presented in Appendix A.6, which describes an in situ bioaccumulation, survival and growth study (Appendix A.6). Comments are provided below relating both to the study as described in Appendix A.6 and to changes as described in the May 6, 1999 meeting. The study protocol describes the possibility of using either growth or glycogen as endpoints in the bioaccumulation study. The significance of these assessment endpoints to the structure and function of aquatic invertebrate populations and communities is not evaluated, and the rationale for using mussels as a component of the food chain study is unclear. Further, the proposed study design is incomplete. The protocol raises, but does not resolve, a number of issues critical to study design, including the designation of reference areas, study duration and the selection of measurement endpoints. When a new protocol is developed for this study, GE requests that copies be provided. Specific comments on the study design described in Appendix A.6 are provided below.

First, the proposed assessment endpoints, survival, growth and glycogen levels, do not meet the selection criteria set forth in Section 7.2.2.1 of the work plan (i.e., they do not “assist in identifying the ecological structure and function at the site”) because they do not address potential effects to subpopulations, populations or communities. Therefore, the assessment endpoints are inconsistent with EPA’s draft risk management guidance and the MCP (see Section 2.4). Survival, growth and glycogen levels are generalized responses and are only appropriate as measurement endpoints to the degree that they can be related to higher level assessment endpoints. However, because invertebrates are relatively insensitive to PCBs, and growth is an insensitive endpoint for PCBs, any observed changes in growth, mortality or glycogen levels are likely to be due to other confounding factors. Unfortunately, as described below, the study design is not sufficiently robust to distinguish other factors (e.g., statistical design, problems with control sites, limited measures to account for other factors) from those effects due to PCBs in sediment. As a result, the use of these parameters as assessment endpoints will lead to an overly conservative estimate of risk to aquatic invertebrate populations and communities in the Housatonic River.

Second, the evaluation of appropriate reference stations is critical to the assessment of natural variability in mussel growth and survival that may be seen in the Housatonic River (see Section 2.10). As currently designed in the study, many variables that can affect growth are not adequately accounted for in the analysis, including water temperature, turbidity, and food supply. Although there will
always be some natural variability between deployment stations due to these factors, it is not possible to quantify this variability adequately without a sufficient number of reference areas. Given these limitations, there is greater likelihood that only weak statistical associations will be used to support claims of PCB risks. The authors state that regression and correlation analysis will be used to determine significant differences among several factors (App. A-6, p. 8). While regression analyses are useful in evaluating associations between measured parameters, it is not clear how other factors that can affect mussel growth will be included in the correlation analysis. All components that will be used in the correlation analysis should be made explicit.

Third, the designation of study locations is incomplete in the work plan. Specific station locations for the caged mussel study are not provided, nor is the tributary that is to be the source of the transplanted mussels specified. Instead, the proposed study design simply states that mussels collected from a "clean unspecified tributary" are to be placed in cages and transplanted to a series of stations, from upstream of the GE facility to downstream of Woods Pond. The only detail provided is that stations are to be selected based on habitat suitability and comparability (App. A.6, p. 2). Based on this limited information, any results that show adverse effects in the sampling locations relative to the control will be difficult to interpret.

Fourth, the authors indicate that the proposed exposure duration for the caged mussel study (i.e., 84 days) may not be sufficient for measurable changes in mussel size to occur (App. A.6, p. 4). This basic issue should have been resolved before finalizing the study design. The proposed size class (55 – 75 mm) may be too broad to allow for distinguishing 2 - 3 mm changes in growth over the duration of the study. Diminishing growth rates are expected as mussels get larger; therefore, it is critical to limit the size class (e.g., to 55 – 60 mm) to adequately differentiate changes in growth. The authors indicate that, if measurable changes in growth do not occur over this period, glycogen levels may be used as an endpoint instead. However, they also acknowledge that glycogen levels may change in response to environmental variables that are unrelated to chemical contamination (e.g., temperature, nutritional status, reproductive status). There is no indication as to how these confounding factors will be addressed in the study design. Moreover, there are no provisions in the study design to determine if statistically significant changes in glycogen levels are relevant to the function of individual mussels let alone the sustainability of subpopulations, populations or communities.

Based on the May 6, 1999 meeting, it appears that glycogen levels will be used instead of growth as an endpoint. However, as previously discussed, the use of a physiological parameter such as glycogen
levels as an endpoint is not a useful measure because there is no connection between any discernible changes in glycogen levels and the population or community.

Fifth, the authors are unsure if the exposure period will be sufficient for the concentrations of COPCs in mussel tissues to achieve steady state. Moreover, they provide no information as to how they will determine if steady state has been achieved, further reducing the ability to accurately quantify either bioaccumulation or growth. Given these problems, statistical comparisons of bioaccumulation, comparisons between sites, and statistical evaluations of exposure-response relationships will likely have little meaning.

Sixth, the authors acknowledge that the use of suspended racks (i.e., no contact with the sediment) with mussels that normally burrow in the sediments may alter the mussels normal feeding behavior and consequently may affect growth and physiological condition (e.g., glycogen content). Deployment in mesh bags may be more effective.

Overall, the proposed study will not provide useful information for use in the ERA, nor will it contribute to the evaluation of remedial alternatives.

3.8 Amphibians and Reptiles

The work plan describes four studies related to reptiles and amphibians, including use surveys, a study of amphibian reproductive success in vernal pools, frog reproductive toxicity testing, and bullfrog and snapping turtle tissue sampling and analysis. Specific concerns related to each of these studies are described in the following sections.

3.8.1 Reptile and Amphibian Use Surveys

Numerous limitations are associated with the reptile and amphibian use surveys (Appendix A.7). Like several other study designs presented in the work plan, little detail is presented on data interpretation, how the various measurement endpoints will be related to the selected assessment endpoints during risk characterization, reference areas, and the study's statistical design. There is also the potential for confounding factors limiting the interpretability and usefulness of the study. Because of the limitations detailed in the following comments, the results of the reptile and amphibian use survey
(which has already been conducted) should be qualified and given little if any weight in risk characterization.

Although not explicitly stated in the work plan, it is our understanding, based on discussions with EPA during the May 6, 1999 meeting, that the reptile and amphibian surveys were principally conducted in order to provide background data for use in designing subsequent studies and in developing the problem formulation. If the survey results are also intended to evaluate risk, it would be necessary to specify one or more points of comparison, such as internal referencing, reference areas, and/or literature data. The protocol does not explicitly state which, if any, referencing method was employed in this survey. To the extent that reference areas were evaluated, they should be named, so that their appropriateness and applicability can be evaluated. Given the sensitivity of reptiles and amphibians to subtle differences in habitats, the selection of reference areas can markedly influence the success and interpretation of the study.

A variety of factors may influence amphibian reproduction, including habitat attributes, traditional water quality measures, and the presence of predators. The protocol does not specify whether such factors were quantified, or whether correlation and regression analyses will be conducted on these factors. Measures of pH, DO, alkalinity, and hardness should be obtained from each survey location. If pH values less than 6.0 are observed, the potential for aluminum toxicity should be evaluated.

Several omissions and/or limitations relate to the statistical design of the survey. First, the data analysis section fails to describe any quantitative, let alone statistical, method of data analysis. While much of the information generated by the survey appears to be qualitative, statistical methods are available to address presence/absence data and catch per unit effort data. Second, no information is provided regarding the statistical basis for the study design, such as the rationale for funnel trapping at 15-20 pools and setting ten traps per pool. Third, the locations of the pit trap arrays should be identified within the protocol. The number of arrays and the statistical basis for selecting the number of arrays are not discussed. Fourth, collecting specimens for PCB analysis based on incidental mortality may not yield a representative cross-section of species, age classes, genders, or PCB concentrations. The study design should be modified to ensure representative and statistically-based sampling of amphibians for bioaccumulation analyses. Finally, samples should not be composited; discrete samples of the same species and size will yield more useful information.
Given the deficiencies in the protocol for the reptile and amphibian use survey, the fact that it has already been completed, and the inclusion of other more rigorous study designs to evaluate amphibian reproduction, the reptile and amphibian use survey should be qualified and given little if any weight in risk characterization. It is not clear that this survey will yield interpretable risk-related results.

3.8.2 Study of Amphibian Reproductive Success in Vernal Pools

The study design for evaluating amphibian reproduction in vernal pools (Appendix A.8) shares many of the same limitations discussed above with respect to the reptile and amphibian use survey design, such as those related to objectives, statistical sampling design, and selection of specimens for tissue analyses. Other limitations of the study design relate to use of data on aquatic macroinvertebrates, duration of monitoring period, and use of length/weight relationships as indicators of health status.

As with other studies, the actual objective of the in situ reproduction study is not clearly stated. If the proposed study is intended for use in characterizing ecological risk, then points of comparison (e.g., internal referencing, reference areas, or literature values) must be specified in the protocol. (Based on observations in the field and discussions during the May 6, 1999 meeting, we understand that internal referencing is being used.) Because amphibian populations exhibit extreme variability in reproductive characteristics, comparability of habitat and water quality among vernal pools is critical. Although not specified in the work plan, we understand from field observations and discussions during the May 6, 1999 meeting that potentially confounding factors (e.g., pH, alkalinity, hardness) are being quantified at all vernal pools; this should help elucidate causation, provided those factors are evaluated using appropriate statistical techniques. The protocol should be revised to specify the methods by which such variables will be measured and evaluated.

The basis for the sampling design is also insufficiently documented. The statistical or other technical basis for studying five pools should be provided. Additionally, collecting adult specimens for PCB analysis based on incidental mortality, as proposed, would not have yielded a representative cross-section of species, age classes, genders, PCB concentrations, etc. However, based on the May 6 meeting, we understand that no adults have been collected as part of this study because there has been no incidental mortality. Also, because the study design calls for marking individuals, sufficient
information may be generated to report estimated population sizes for each species. As part of risk characterization, such data should be given greater weight than the results of the use survey.

The basis for the analytical approach is also insufficiently documented and appears inadequate in several respects. First, the work plan states that "effects will be assessed using stressor-response curves and no-effect levels." The method for developing and interpreting the stressor-response curves and no-effect levels should be described in greater detail. Second, the use of a 0.10 alpha level is not generally acceptable for scientific studies. The majority of technical journals that report the results of herpetological studies require the use of an alpha level of 0.05, as do other studies proposed as part of the Housatonic River ERA (e.g., tree swallow nest box study). Third, the technical basis for the proposed 10-day monitoring period is not provided. The monitoring period should be proportional to the development period for each species targeted for study. Time to metamorphosis should be considered as a possible endpoint for the study. Fourth, the study design should also describe the application of data collected for aquatic macroinvertebrates that are inadvertently trapped in the funnel traps.

Finally, the use of length/weight relationships as an indicator of the health status of amphibian populations is entirely theoretical; empirical studies are lacking to document such relationships. Most amphibian studies that evaluate length and weight have been concerned with documenting patterns of sexual dimorphism and the evolutionary implications of such patterns. While occasional reports of abnormal (e.g., emaciated) specimens occur in the scientific literature, an empirical analysis of length/weight relationships as related to health status of the amphibian species selected in the proposed study is not available. In addition to various biotic and abiotic environmental factors that may affect length/weight relationships, seasonal fluctuations also occur and are usually more pronounced in female amphibians as a result of physiological and behavioral characteristics associated with pre- and post breeding condition. The specific technical justification for the use of this specific measurement endpoint should be detailed.

In light of the numerous deficiencies in the protocol for evaluating amphibian reproductive success in vernal pools, as well as the inclusion in the work plan of a far more rigorous study design to evaluate reproduction in leopard frogs, the protocol and study design for the in situ evaluation of frog reproduction in vernal pools should be revised. Based on the existing protocol, this study will not yield interpretable results for risk assessment purposes.

Indeed, this may be the objective of marking individuals, but that is not stated explicitly in the work plan.
3.8.3 *Frog Reproductive Toxicity Testing*

The frog reproductive toxicity testing protocol described in Draft Version 3 of Appendix A.9 is of markedly higher quality than many of the other study designs presented in the work plan. Much of that protocol's text was drawn directly from a study design developed and presented by ChemRisk (1997). Indeed, due to that source, there remain vestiges of the original text that are no longer applicable to this protocol. For example, page 5 of Appendix A.9 notes that safety training and medical monitoring requirements are consistent among all protocols for the field studies, but those requirements are not specified anywhere else in the work plan. Likewise, before there has been any discussion of candidate reference areas, page 5 notes that, "following the agencies' approval of this protocol, these and other potential reference areas will be evaluated in greater detail." It should also be noted that, given the date of preparation of Version 3, the reference areas had likely already been selected when the protocol was re-issued. As such, it would be appropriate to name the selected reference areas within the protocol.

3.8.4 *Bullfrog and Snapping Turtle Collection and Tissue Analyses*

Bullfrogs and snapping turtles will be sampled for tissue analyses, as described in Appendices B.11 and B.12, respectively. The analytical results will be applied to both the human health risk assessment and the ERA. While we understand from discussions held on May 6, 1999 that the primary objective of this effort is in support of the human health risk assessment, limitations of the study design may preclude the resultant data from being useful in the ERA. First, in order to yield data that are representative of the diets of ecological receptors, the work plan should identify specific receptor species, and characterize their preferred diets with respect to species and size. The sampling plan should then be modified to target the sizes and species of preferred prey items for the selected receptor species. Absent such detail, it is not clear that this study will provide useable, interpretable, or necessary data for the ERA, particularly when more representative tissue analyses will be conducted as part of the other studies on amphibians. Based on discussions held on May 6, it is apparent that EPA realizes that tissue data generated by the other amphibian studies will likely yield data that are more representative of prey consumed by ecological receptors. The work plan should be revised to specify the criteria that will be considered in judging the appropriateness of tissue data for use in TQ calculations and to acknowledge that all available tissue data (not just bullfrogs and snapping turtles) will be evaluated for potential use in TQ calculations.
3.9  Fish

The work plan describes two major studies related to Housatonic River fish. The first study involves the comparison of fish tissue concentrations collected on the Housatonic River with MATCs from the literature. The second, the fish health and toxicity testing study (Appendix A.10), has two phases, (a) laboratory egg injection studies and (b) laboratory rearing of field-collected fish eggs. Although not described as a distinct study, a third aspect of this study is an evaluation of biochemical indices of endocrine disruption in adult fish. Specific issues related to each of these studies are described in the following sections.

3.9.1  Comparison of Fish Tissue Concentrations with Established Effects Levels

The ERA work plan indicates that potential effects on fish will be evaluated by comparing the concentrations of COPC in fish tissues to MATC levels found in the published literature (p. 7-45 and Table 7-1). The tissue PCB data upon which the MATC comparison will be based were collected in 1998; therefore, comments and/or concerns identified by GE (and others) cannot be implemented, unless the entire study is either redone or omitted from the ERA. The MATC approach is an indirect method of evaluating survival, growth and mortality and is inherently conservative. Furthermore, it relies upon generic information that does not account for site-specific factors but does require the use of uncertainty factors and extrapolation. The comparison of tissue concentrations with MATCs to evaluate potential effects is constrained by several issues, including comparison with studies that have different exposure scenarios, different COPCs and species different from those found in the Housatonic River (see Section 2.7). Interspecies extrapolations are likely to be required for most fish found in the Housatonic River; however, given that the sensitivities of fish species to PCBs vary by several orders of magnitude, these results may be highly inaccurate. Given these uncertainties, a conservative approach is likely to be taken resulting in a significant potential to overestimate risk to fish.

Adequate justification for a field collection and analysis of this size is not presented. Based on the May 6, 1999 meeting, we understand that the study was designed to serve several purposes (e.g., human health risk assessment, modeling, food chain exposures). However, there is no clear indication of how these data will be specifically used in the MATC comparison (e.g., which species, locations, analyses) or if comparisons to the MATC will be carried out for COPCs other than PCBs. The study
collection and the associated chemical analyses [PCBs (Total and Aroclor) N=1007, dioxins/furans and organochlorine pesticides (n=768)] (p. 5-81) represent a major expenditure, but do not provide data of comparable value. Although there is relatively little congener-specific toxicity data available in the literature, total PCBs have been found to better correlate with effects in fish than congeners, as discussed in Section 2.14 above and explained further in Section 3.9.2.1. Therefore, conducting congener-specific analyses on such a large scale is unwarranted. Several years of PCB data on young-of-year and adults from the Housatonic River have been collected and would have provided an adequate basis for comparisons with MATCs.

Considering the limitations associated with this approach, EPA should not rely on MATCs and this study should be omitted from the work plan. If MATCs are to be used despite these limitations, selection criteria and specific values should be specified and justified in the work plan. At the May 6, 1999 meeting, EPA explained that it would not consider the comparison of fish tissue concentrations to MATCs as a strong line of evidence, but included that comparison as an endpoint only because the fish tissue data are available. If this is the case, it should be stated in the work plan. This is another example of the lack of clearly specified methods for integrating multiple lines of evidence, as discussed in General Comment 2.3.

3.9.2 Fish Health and Toxicity Testing

The “Fish Health and Toxicity Testing” section of the work plan includes several studies that are presented in detail in Appendix A.10. These studies include: (a) a laboratory injection study in which eggs from several fish species will be injected with concentrations of an organic extract of tissues from fish collected from the Housatonic River; (b) a reproduction study in which larvae from eggs collected from Housatonic River fish will be raised in the laboratory; and (c) a study that will measure several physiological and biochemical parameters that appear to focus on endocrine disruption in adult fish collected from the Housatonic River. Neither the laboratory injection study nor the endocrine study should be carried out for reasons that are detailed below. The reproduction study on eggs collected from feral fish is a valuable study, but the study design is incomplete. Specific issues with each study are provided below.
3.9.2.1 Phase One – Laboratory Egg Injection Studies

The relevance of a laboratory egg injection study to the exposure of feral fish to PCBs through maternal transfer is unsubstantiated and is clearly not, as stated, “[t]he most direct measure of embryo toxicity of contaminants present in fish from the Housatonic River…” (App. A.10, p. 2). This technique for evaluating fish reproductive potential is experimental in nature (see Section 2.15). Furthermore, the work plan proposes a separate study that directly assesses the reproductive potential of fish from the Housatonic River. Based on these concerns, elaborated upon below, this study should not be carried out. Nonetheless, a number of detailed comments on the study, including the development of the organic extract, the method of exposure, species selection, and the use of the TEF/TEQ approach are discussed below.

It is inappropriate to use development and maturation assessment endpoints for fish (Table 7-1). These proposed endpoints do not meet the selection criteria for assessment endpoints as set forth in Section 7.2.2.1 of the work plan (e.g., they do not “assist in identifying the ecological structure and function at the site”), in that they do not deal with subpopulations, populations or communities, and thus are inconsistent with EPA’s draft risk management guidance and the MCP (see Section 2.4). To be consistent with EPA guidance, the assessment endpoints should focus on potential effects to populations and communities. The use of the proposed parameters as assessment endpoints will result in an overly conservative estimate of risk to fish populations and communities in the Housatonic River. As indicated in the following paragraphs, this problem is exacerbated by employing an exposure method (egg injection) that does not reflect maternal transfer of PCBs and laboratory reared fish species, several of which are not native to the Housatonic River.

GE has a number of concerns regarding the development of an organic extract, a critical component of the egg injection study. The source species for the organic extracts to be used in the egg injections have not been specified nor have the criteria to be used for selecting these species been established. An extensive fish collection was conducted in 1998. Aroclor and congener/homolog analyses were conducted on a total of 1,007 samples from eight species of fish from seven locations, including two reference sites (Table 5.5-1). Given that an extensive fish collection effort has already been completed, it is inappropriate that the source species for the development of the organic extract is not yet designated.
The work plan proposes the injection of organic extracts ["hopefully the same species from each study area" (App. A.10, p. 4)] into three different species; however, the injection of an extract from one species into another is experimental and does not mimic maternal PCB transfer to eggs. In addition, the specific mix of congeners in fish tissues and the resulting toxicity, will vary depending upon the feeding strategy and trophic level of the fish species. The use of the same extract for the three proposed test species is inappropriate. Moreover, the use of multiple species as sources for the organic extract would equate to different treatments, so comparisons between study areas would not be possible. A final issue is that the chemical extracts are to be obtained from composites of whole fish. Because the mix of congeners varies among tissues, it is likely that the congeners taken from the whole fish extract will not reflect the mix of congeners (or the toxicity of PCBs) in eggs of fish (Mac et al., 1993). As proposed in the work plan, there is no method to calibrate the range of injected doses from the organic extract to assure that they reflect the range of PCBs found in eggs of fish from the Housatonic River. In the May 6, 1999 meeting, EPA indicated that range-finding exercises would be carried out to determine the appropriate range of congener concentrations. No study design for or results from such a calibration have been made available to GE.

The timeline of this study plan indicates that the development of the organic extract for egg injections will take place from March 1999 to August 1999, and chemical characterization of the fish eggs will not occur until June 1999 through December 1999 (App. A.10, p. 20). This is an inappropriate sequence as it is absolutely critical to characterize congeners in eggs prior to the development of the extract in order to ensure the selection of a species with a congener mix that most resembles that found in eggs. As a result of these problems, it will not be possible to extrapolate from any effects that may be found in this study to fish residing in the Housatonic River.

The choice of test species from which eggs will be obtained for injection is also inappropriate. The proposed species include hatchery raised rainbow trout, fathead minnows, and bass or bluegill. Neither the rainbow trout nor fathead minnow are found in the reaches of Housatonic River under study. In the May 6, 1999 meeting, the EPA indicated that it believes rainbow trout are a representative species of the Housatonic River system. However, rainbow trout do not and have not naturally occurred in the Housatonic River. The population of brown trout that currently inhabits the reach upstream from Dalton Pond dam and throughout a large portion of the Housatonic River in Connecticut is maintained by annual stocking. No natural reproduction is known to occur anywhere throughout the system. A cold water fish such as the rainbow trout is not a functioning component of this warm water river system (i.e., they do not complete their life cycle within the system) and, therefore, is an inappropriate
species to select as a representative indigenous species for a reproductive study. Moreover, of the fish species that have been tested to date, rainbow trout is the most sensitive to PCBs, and a comparison of effects with Housatonic River fish is likely to overestimate risk.

The selection of adequate and sufficient reference sites is critical in order to bracket the range of habitat conditions to which fish in the Housatonic River are exposed (see Section 2.10). This study proposes use of only one reference site, which is neither adequate nor representative of the full range of background conditions. It is not clear why only one reference site was selected since the fish collection study conducted in 1998 used two reference sites, which was minimal given the varied habitat and conditions under consideration. In order to adequately reflect the range of conditions in the Housatonic River, several more reference sites should be selected.

The work plan proposes to assess the exposure concentration of the extracts that are to be injected into the fish using the TEQ/TEF approach. This is unfounded because it assumes an additive model of toxicity. Although the TEQ/TEF concept provides a means of evaluating PCB exposure on a congener-specific basis, this approach is based on a number of important assumptions and has limitations related to these assumptions, especially for PCBs. The TEQ/TEF approach assumes that: (a) the toxicities of individual PCBs are additive when combined in mixtures; (b) no variability occurs in sensitivities between endpoints and within broad species groups; (c) effects of PCBs are due solely to the presence of dioxin-like congeners; and (d) the dose-response curve for 2,3,7,8-TCDD is parallel to that for individual congeners. Exceptions to all of the assumptions have been reported in the literature for fish (Safe, 1994; WHO, 1997; Pohjanvirta et al., 1995; Putzrah, 1997; Starr et al., 1997). For example, TEFs or the relative potency of individual congeners can vary considerably by species, and therefore TEFs developed for one species may not be applicable for other species. Results of laboratory studies also suggest that certain PCB congeners that act through mechanisms unrelated to the Ah receptor may also elicit biochemical and toxic responses (Safe, 1994). Effects not mediated by the Ah receptor are not considered in the TEQ/TEF approach (see also Section 2.14 above).

The protocol for this study appears to be experimental and requires significant refinement and validation (see Section 2.15). The ERA for the Housatonic River is not the appropriate venue to achieve these research and method developments. It is GE's understanding, based on the May 6, 1999 meeting, that it would receive modified protocols for the injection study and that it would be free to contact the principal investigator in order to clarify and discuss these issues. Neither event has occurred as of this writing.
In summary, the egg injection study described in Appendix A.10 is a flawed, cumbersome and indirect approach for evaluating the effects of COPCs in the Housatonic River on early life stages of fish. It will likely yield ambiguous results that will substantially overestimate the potential risk to fish in the Housatonic River. This study should not be conducted.

3.9.2.2 Phase Two - Laboratory Rearing of Field-Collected Fish Eggs

Phase Two of the fish health and toxicity study has many of the same limitations as Phase One with respect to the relevance of the assessment endpoints to populations, species selection and reference areas. The study design is vague, incomplete and flawed such that it will limit the extent to which the results can be interpreted. However, if properly designed and implemented, this study could well provide a coherent assessment of the potential for adverse effects due to PCB accumulation in tissues of fish from the Housatonic River. Further issues that need to be addressed include the statistical sampling design, vagueness in the experimental design, use of the laboratory study to “calibrate” results from the field collected fish, and insufficient reference sites. These issues are discussed in detail below.

The sampling plan is exceedingly vague; therefore, it is not possible to assess the level of statistical rigor it will support. Appendix A.10 does not specifically indicate how many adult females or males will be used in the whole study or at each of the sampling locations, nor is the number of replicate samples of eggs per female specified. Between-female variance decreases and random error increases with each successive measure of reproductive success from spawning through hatching (Spies and Rice, 1988). As a result, in order to detect differences in embryological success or proportions of normal larvae between sites, it is important to carefully determine how many fish need to be spawned.

There is not enough detail in this experimental design to ascertain whether the eggs from different adult females will be kept separate so that any observed embryo-larval effects can be linked to maternal COPC concentrations. A study design should be proposed that would track PCB concentrations in corresponding eggs and adults so that correlations with adult exposure can be evaluated. Furthermore, the study design does not specify how many fish or egg samples will be analyzed.
The work plan specifies that the egg injection studies will “serve as a standard curve for calibrating any effects observed in the field-collected fish” (App. A.10, p. 17). This indicates that the egg injection study will serve as the primary tool for evaluating fish embryo larval toxicity. The eggs from fish from the Housatonic River, which provide site-specific information on fish health, will be used only as a “validation exercise” (App. A.10, p. 16). This is clearly an inappropriate way to evaluate potential effects in the field and serves more as an exercise in evaluating the appropriateness of the experimental egg injection study. A similar unsubstantiated claim is that “…the appropriate selection of endpoints in the field studies will allow the characterization of any causal linkages among contaminant exposure and adverse effects through a comparison with the laboratory based results of these studies” (App. A.10, p. 3).

Only one reference site is proposed for this study. This is inadequate and will not represent the range of habitat conditions to which fish in the Housatonic River are exposed and will not provide a proper reflection of natural variability (see Section 2.10). It is not clear why only one reference site was selected since the fish collection conducted in 1998 used two reference sites, which is considered minimal. In order to adequately account for the conditions of the Housatonic River that are under consideration, several more reference sites should be selected.

Although GE raised several of these concerns with EPA in a conference call before the fish reproduction study was commenced, EPA refused to make any changes in the study. EPA’s failure to include additional reference sites and sites along the Housatonic River in this study is particularly unwarranted. EPA indicated at the May 6, 1999 meeting that one of the main reasons for conducting the egg injection study was to address concerns regarding the development of an adequate dose-response curve from the reproduction study based on field collected fish. The inclusion of additional sites along the Housatonic River and additional reference sites in the reproduction study would have been an important step to improve the resolution of the dose-response relationship. To the extent that EPA can revise this study based on the above comments, we urge it to do so.

3.9.2.3 Evaluation of Endocrine Disruption

The ERA work plan calls for the evaluation of several hormonal/biochemical parameters in adult fish in conjunction with COPC analysis in fish tissues. Endocrine studies, although increasingly common, have yet to demonstrate a link between biochemical changes and effects on individuals, much less effects at the population level (see Section 2.15). The parameters to be evaluated include EROD
induction, estrogen/testosterone ratios, and vitellogenin synthesis. EROD induction is an indirect measure of Ah receptor-mediated enzyme induction and is used as a measure of exposure to chemicals that act via the Ah receptors (e.g., coplanar PCBs, dioxins etc.). The latter two measures are included to evaluate the potential for endocrine disruption. However, a recent study with rainbow trout showed no relationship between high exposures to a planar PCB congener and vitellogenin synthesis (Donohoe et al., 1999). Moreover, the study has no provisions for determining if any changes in these biochemical parameters will result in reduced fecundity or larval viability, let alone population level effects. The proposed study will not provide data that will help to evaluate potential risk to fish in the Housatonic River and should not be conducted.

Furthermore, the EPA should not attempt to use the Housatonic ERA to conduct research designed to overcome these method limitations. The Superfund process is not intended to serve as a research project for methods development and validation. In any event, the studies in this work plan do not appear to be sufficiently focused or statistically rigorous to support such an effort.

3.10 Earthworms

Earthworms are the only terrestrial invertebrates to be evaluated in this work plan. The use of this study as the sole representation of terrestrial invertebrates is insufficient and inappropriate. The proposed study is deficient both in its evaluation of earthworm assessment endpoints and in the estimate of contributions made by terrestrial invertebrates to exposure of higher trophic levels via the food chain. As with many of the proposed studies, the assessment endpoints are inappropriate, the designation of target and reference locations is vague, and it is not likely that the sampling design will support a rigorous statistical evaluation. Additional problems with this study include the use of inappropriate methods for evaluating PCB concentrations in earthworms, the use of the MATC approach, and a proposal to derive an exposure-response relationship unsupported by the study design.

Although survival and growth are typically used as measurement endpoints, their use as assessment endpoints is not supported by EPA guidance because they only address effects at the level of the individual. As a result, the use of these assessment endpoints does not meet the selection criteria set forth in Section 7.2.2.1 of the work plan (e.g., they do not “assist in identifying the ecological structure and function at the site”) and is inconsistent with EPA’s (1998a) draft risk management guidance, which clearly states that populations are the appropriate level of ecological organization for assessment in this context (see Section 2.4). Furthermore, the study design proposes an evaluation of
effects to earthworms through a comparison of concentrations of PCBs found in earthworms with toxicity values found in the literature. This approach does not allow for a rigorous evaluation because it does not reflect site-specific conditions and requires the use of uncertainty factors that increase conservatism and reduce realism. During the May 6, 1999 meeting, EPA acknowledged that the use of MATCs derived from the literature provides a weak line of evidence. GE agrees and expects that little weight will be given to the MATCs as a line of evidence. A properly designed study would evaluate survival and growth using readily available, established and inexpensive laboratory bioassays to evaluate risk to earthworms directly. As currently proposed, the use of these parameters as assessment endpoints will result in an overly conservative estimate of risk to earthworms and will be insufficient to estimate risk to other soil invertebrate populations and communities in the Housatonic River floodplain.

In addition to the stated assessment endpoints, analytical data collected on earthworms will be used in the food chain modeling to represent dietary contributions of terrestrial invertebrates to insectivorous bird and mammal species (p. 5-75). The use of earthworms as the sole representatives of the invertebrate food chain is inappropriate because their high lipid content will lead to an overestimation of dietary dose up the food chain. Instead, a broader variety of terrestrial invertebrate species should be sampled to more accurately characterize the dietary composition of carnivores and to provide a better estimate of PCB contribution to the food chain. If no changes are made to the study design, the implicit limitations of sampling only one potential prey species for this guild should be discussed. Based on the May 6, 1999 meeting, GE understands that this protocol may be modified to include more terrestrial invertebrates.

The sampling design is vague and incomplete. The work plan provides inconsistent and overly general criteria for the selection of sampling sites and neither the number nor the location of target and reference sampling sites nor the number of samples to be taken at each site has been finalized. The authors acknowledge that they have not provided a statistical analysis of sample size required to detect statistically significant differences in concentrations of PCBs; however, it is not clear if they have actually completed such an analysis (App. B.10, p. 1). The specifics and the rationale behind this sampling design should be more explicitly presented.

The work plan proposes to use data collected on earthworms in two distinct ways: (a) to estimate the dose to the worm for comparison with literature derived toxicity thresholds and (b) to estimate the delivered dose for food chain modeling. Directly compositing and measuring PCB concentrations in
earthworms is appropriate to estimate delivered dose for this component of the food chain model (but not for all soil invertebrates). However, the use of this same method to estimate the dose to the worm itself would result in a gross overestimation of exposure because much of the mass of COPCs is associated with the gut content rather than the tissue of the worm itself. To address this, the earthworms should undergo depuration in order to empty the gut of contaminated soil that is not representative of the tissue concentration of PCBs. It is critical that any literature comparisons are for studies that have dealt similarly with depuration. These issues should be addressed, the methodology should be adjusted accordingly, and additional detail should be provided in the work plan.

Although the work plan implies that attempts will be made to develop an exposure-response curve and a no-effect level for earthworms (p. 7-45), we understand, based on the May 6, 1999 meeting, that no stressor-response curve will in fact be developed. Indeed, the only evaluation of responses (effects) described in the work plan is based on comparisons to literature-based no-effects levels. Limitations of comparing tissue concentrations with toxicity thresholds reported in the literature include: varying species sensitivity; comparisons to earthworm toxicity tests using a variety of test methods, some of which do not mimic exposure in the field; and comparisons with studies using endpoints which are not directly relevant to the endpoints of interest. As previously discussed, a relatively inexpensive and straightforward bioassay with soil collected from the site would provide a direct measure of stressor-response. Furthermore, the proposed sampling design is too vague to determine if sampling will encompass an adequate range of PCB concentrations to be used in estimating exposure for such a study.

Overall, this is a flawed and incomplete study design. As proposed, it will not adequately address toxicity to earthworms nor will it accurately characterize the contribution of terrestrial invertebrate species to the dietary composition of insectivorous birds. The design of this study is appropriate for use in a food chain model (if modified to include other terrestrial invertebrates), but it should not be used to assess toxicity to soil invertebrates. This study needs to be substantially revised to include the evaluation of PCB concentrations in a range of terrestrial invertebrates and to address the deficiencies described above. It is GE’s understanding that modified protocols for the terrestrial invertebrate study will be made available in a timely fashion and that it will be possible to directly contact the principal investigator in order to clarify and discuss these issues.
3.11  Tree Swallows

Limitations of the proposed tree swallow reproduction study (Appendix A.12) relate to the selected reference and sampling areas, parameters selected for evaluation, insufficient level of detail, and inconsistencies. Also, as discussed in Section 2.1, because the first year of this two-year study has already been completed and the second year is underway, weaknesses and deficiencies identified by GE (and others) are unlikely to be remedied in the study design. Consequently, some of the data generated from this study may not be useable in the ERA.

A major limitation of the tree swallow reproduction study design relates to the selected reference area. Given the topography, habitat, and human development in the vicinity of candidate reference areas (nearby tributaries, the West Branch of the Housatonic, and the Housatonic River upstream of the facility), the selected reference areas may not be appropriate. The study design should have included a pre-screening step for identifying suitable reference areas using predefined criteria. Criteria that should have been considered in selecting the reference areas include: proximity to a river, distance from contaminated sediments and human development, and quality of vegetative habitat as a tree swallow breeding area.

Additionally, the reference area described in the work plan is located too close to the study area. Male tree swallows may wander 5 km from the nest site (Robertson et al., 1992) and minimum foraging distances in a wetland were estimated by Blancher and McNichol (1988) to be 2.2 km. The identified reference area is within 2.4 km of contaminated sections of the Housatonic River and swallows from the proposed reference area could forage within the study area. Assuming that more comparable habitat can be found more distant from the source, the reference area should be relocated.

In addition to problems with the selected reference area, sampling areas along the Housatonic River are as little as 1.4 km apart. This concentrated grouping of sampling areas only allows for sampling at the upper end of the concentration gradient. Distribution of sampling areas across a longer stretch of the Housatonic River would show the effects of exposure to variable concentrations of PCBs at various points in the concentration spectrum. Greater spacing of sampling sites along the Housatonic River would eliminate possible overlap of home ranges and result in measurement of tree swallow variables over a greater range of exposure conditions.
Several necessary parameter measurements are absent from the work plan, including breeding female age and adult plumage stage. All of these parameters should be included in a study that aims to determine if tree swallows are affected by exposure to PCBs. Of the parameters selected for evaluation, several modifications are necessary.

First, the age of breeding females can influence reproductive success. Sub-adult female tree swallows are known to be less successful than older, more experienced females (Wheelright and Shultz, 1994). Thus, the proportion of sub-adults and adults in the sampled tree swallow population on the Housatonic River is important for determining expected or average reproductive success. It is recommended that females be aged using two methods. Plumage coloration is often used to age birds; however, it should not be used alone because chemical exposure is suspected of causing premature development of female tree swallow plumage (Secord and McCarty, 1997). Therefore, skull pneumatization should also be used to verify the age of breeding females (Proctor and Lynch, 1993).

Second, pippers should not be sampled from every nest. In comparison, tree swallow studies conducted on the Hudson River did not collect eggs and fledglings from every nest (Secord and McCarty, 1997; McCarty and Secord, 1999). Rather, a representative subset of nests should be left untouched to allow accurate estimates of hatching and fledging rates and to allow accurate implementation of the Mayfield method. Appendix A.12 (p. 6) states that the Mayfield method will be used to calculate reproductive success. The Mayfield method divides nest success (reproductive success) into five stages: survival during nest-building, survival during egg-laying, survival during incubation, the number of eggs hatched, and survival of young to fledging (Mayfield, 1975). Removal of eggs or nestlings from a nest may affect the survival of the remaining young, thereby skewing the calculation of the last three stages of Mayfield's measures of nest success.

Third, the distance between nest boxes should be consistent between sites to eliminate variation in nest density effects. Nest density can affect breeding behavior and reproductive success of tree swallows (Rendell and Robertson, 1989). Varying the distance between nest boxes by as much as 10 meters could cause variations in reproductive success at individual nest boxes. As previously discussed, greater spacing of sampling sites along the Housatonic River would also eliminate possible overlap of home ranges and result in measurement over a greater range of exposure conditions. An historic study on the life history of tree swallows by Kuerzi (1941) conducted on the Housatonic River prior to the deposition of the vast majority of PCBs in the river warrants consideration as a point of comparison. Kuerzi (1941) measured many of the same reproductive endpoints that are proposed by Dr. Custer,
potentially offering historic baseline information that avoids many confounding factors (e.g., geographic variation, habitat differences).

Fourth, careful notes should be collected to monitor for effects of parental disturbance on nest abandonment. Tree swallows are quite tolerant of human disturbance. However, if nests are frequently disturbed or adults are extensively handled (e.g., physical removal from the nests), there may be increased nest abandonment.

Fifth, greater detail should be provided regarding the method of ligaturing nestling tree swallows to obtain additional food samples. Ligaturing is mentioned in the work plan as a method to collect additional food samples from nestlings. It is recommended that if ligaturing is performed, cable ties be used as opposed to pipe cleaners or other wires. Mellott and Woods (1993) found that proportionately fewer nestlings died and more food boluses were collected using cable ties compared to pipe cleaners.

Sixth, the study design is not truly blind because pipper, nestling, and food samples will be labeled according to collection location, thereby possibly biasing the laboratory analysts. It is recommended that labels not contain a complete location name. Codes should be created to represent the different sampling areas so that the laboratory technicians cannot be biased during sample analysis. Ways of controlling bias include: numbering samples randomly so that the sample identification number provides no information on its location; processing samples in the laboratory; and ensuring that the person processing samples does not have access to the key that links sample numbers with locations.

In addition, an insufficient level of detail is provided on the actual study methodology. For example, greater detail should be provided regarding proposed methods of data analysis and evaluation, as well as the methods that will be used to judge potential effects. The method of estimating the accumulation rate should be specified. Likewise, more information should be provided regarding the method of chemical analysis, and quality assurance/quality control measures. The selected laboratory's standard operating procedures should be attached to this protocol. While the protocol notes that "prior to any statistical analysis, the data will be verified for accuracy," it is not clear whether this verification process includes formal third-party data validation, and if so, at what tier. Power analysis is not mentioned among the statistical tests that will be conducted.
Finally, there are several inconsistencies in the proposed sediment sampling design for the tree swallow study, based in part on the assumptions that PCB concentrations are greater in Housatonic River sediments than in adjacent backwater areas, and that sample heterogeneity within the 400-m foraging radius is minimal and differences can largely be explained by grain size and organic carbon content. Blasland, Bouck & Lee's sediment sampling results do not consistently support these assumptions (BBL, 1996). Additionally, the work plan is inconsistent in its description of the numbers of co-located sediment samples that will be collected. Table 5.2-2 indicates that 170 sediment samples will be collected, while page 5-31 indicates in two different places that either 141 or 181 sediment samples will be collected.

In conclusion, limitations of the proposed tree swallow reproduction study relate to the selected reference and sampling areas, parameters selected for evaluation, insufficient level of detail, and inconsistencies within the study design. These concerns are made all the more significant because the study has already commenced. Consequently, the resultant data will need to be qualified with respect to uncertainty, and may be only partially useable. The design for this study needs to be substantially revised.

3.12 Other Avian Species

The work plan proposes a number of studies related to other avian species, including collection of waterfowl for tissue analyses and surveys of raptors, forest birds, and marsh and wading birds. While the specific limitations of each of these study designs are detailed below, the surveys share certain limitations related to their objectives. The surveys of raptors, forest birds, and marsh and wading birds do not appear to have been designed to directly answer questions regarding potential ecological risks. Although the work plan does not state so explicitly, the surveys appear to have been conducted to support problem formulation and to select receptors of interest for evaluation in using a literature-based TQ approach (p. 5-87). Rarely (if ever) are such elaborate studies executed in support of such limited objectives. Even so, the receptors of interest that were selected as a result of those surveys should be explicitly identified in the work plan, and the rationale for their selection should be documented. Alternatively, if the surveys will also be used to predict potential ecological risks, the results must be compared to values that would be expected in areas with comparable habitat but without PCBs, either reflecting a reference site or literature data. If predicting ecological risks is

---

1 Forest bird and marsh and wading bird surveys were completed in 1998.
4 A reference area is mentioned parenthetically in the description of the raptor survey provided on page 5-88 of the work plan, but no additional information is provided.
Indeed one of the objectives of the surveys, then the reference areas and/or literature studies should be specified, along with proposed methods of determining whether differences exist and whether those differences are biologically and statistically significant. Indeed, modification of the survey designs to allow their use in predicting risks would likely yield more meaningful results than would be provided by the literature-based TQ approach, in that the surveys are site-specific and could focus on effects at the population and/or community level. In addition to these general observations, limitations of the various studies proposed for other avian species are discussed in the following sections.

3.12.1 Collection of Waterfowl

As discussed in Section 2.1, because waterfowl have already been collected, comments and/or concerns identified by GE (and others) cannot be implemented, unless the entire study is either redone or given little if any weight during risk characterization. In addition, there are numerous limitations associated with the MATC approach, as described in detail in Section 2.7. The primary additional limitations of the protocol for collecting waterfowl relates to insufficient detail describing key elements of the overall methodology, such as reference areas, the basis for MATCs, the statistical sampling design, and methods of sampling and analysis, as discussed below.

In contrast with most other study designs, Appendix A.11 does name and describe the selected reference area. However, additional information regarding the waterfowl reference area should be provided in the work plan, beyond simply describing it as “appropriate” (p. 7-46). For this study design, sampling from a reference area appears to be intended to define a background level of COPCs in duck tissues. Because ducks are migratory and may be exposed to a broad range of COPCs both in their breeding and wintering ranges, a range of background tissue residues should be defined by collecting multiple samples from at least two different reference areas. Additionally, the specific statistical method of comparing study area and reference area waterfowl tissue residues should be specified.

The work plan should specify the criteria that will be used to select MATCs, as well as the specific MATCs that will be compared to duck tissue residues. Defining MATCs during the planning stage of the ERA will ensure that “apples are compared to apples” – i.e., that the tissue analyzed, preparation methods, and analytical methods are comparable – and that the effects level selected is the highest quality and most applicable value available.
Finally, additional information should also be provided regarding the statistical study design. For example, the statistical (or other technical) basis for the intended sample size (20 ducks per species per location) should be specified (p. 5-84). Greater detail should be provided regarding data analysis and evaluation, as well as the methods that will be used to judge potential effects.

3.12.2 Raptor Survey

Flaws in the proposed raptor survey (Appendix A.13) relate to the justification provided for conducting the survey, the survey’s objectives, and inability to provide information regarding potential ecological risks posed by COPCs, insufficient detail regarding the proposed methodologies, and the feasibility of executing the survey successfully so that it generates useful information.

It appears that the basis for initiating the raptor survey was that northern harriers, sharp-shinned hawks, Cooper’s hawks, and long-eared owls were not observed during incidental observations made during the course of field work in 1998. First, even among experienced bird watchers, sighting these four species (or their nests) would be unusual due to their secretive behavior and low density. Second, in contrast with the general references cited in the protocol, according to a breeding bird atlas specific to the state of Massachusetts (Veit and Peterson, 1993), these species are not expected to be observed in western Massachusetts and have not been known to breed in western Massachusetts since the 1950s. In Massachusetts, northern harriers only breed on Cape Cod and the islands. There have been only a few confirmed instances of breeding sharp-shinned hawks in Massachusetts since 1955 and those were generally in the eastern part of the state. Likewise, there have been very few confirmed records of Cooper’s hawks breeding in the state (i.e., not more than one pair statewide in any given year) and those too have generally been limited to the eastern part of the state. Veit and Peterson (1993) also remark that it “is exceedingly difficult to find conclusive evidence of long-eared owls breeding in Massachusetts.” The few records of breeding long-eared owls in the state since 1955 occur in the eastern part of the state. Since Veit and Peterson (1993) is cited in Protocol A.14-2, the authors were clearly aware of this reference and some explanation should be provided as to why it is not relied on here.

The actual goals of the raptor survey appear to contradict the stated objectives. Despite the inclusion of fledge rates as a possible endpoint and mention of a reference area (both of which suggest that the survey will strive to evaluate potential ecological risks), Appendix A.13 states that the survey information will simply be used to help develop food chain models and identify species of
management concern. Such a complex study is not warranted to meet these objectives. However, if the proposed study design were substantially modified (as discussed below), it could offer far more useful, site-specific, and ecologically relevant information regarding potential risks to raptors than would be provided by a generic, literature-based TQ. Regardless, the purpose and application of the survey results should be decided upon and explicitly stated in the work plan, and the protocol should be modified to ensure that the survey will in fact meet those goals. For example, given these species' distribution and low density, a defensible survey design could be developed which yields information on presence/absence and diversity, but would not likely provide meaningful information on reproduction or fledge rates.

Like many of the other study designs, sufficient detail is not provided regarding the specific methodologies to be employed as part of field activities, data analysis, and interpretation. The protocol should specify that the same observer will conduct each replicate visit, and that there will be one designated observer and one designated note-taker. Given variability in observer skill, it is important to use the same observer for each replicate visit. Likewise, without a designated note-taker, observations may be missed while the observer is taking notes. More information should be provided regarding the reference areas surveyed for the raptor study (i.e., beyond the simple fact that a reference area will be surveyed). Given the importance of prey and habitat in determining the presence of various raptor species, multiple reference areas should be surveyed and very closely matched to the study area. Quantitative habitat surveys should be conducted to verify the reference areas' and study areas' comparability. Specific criteria for selecting reference areas should be provided in the protocol. Ordinarily, a protocol that includes evaluation of nestling fledge rates should describe the specific methods to be used; however, because it appears infeasible to measure this endpoint, it should be omitted from the methodology altogether. The protocol should include sample field survey forms. Specific methods of data analysis and interpretation should be detailed.

As currently proposed, limitations in the proposed raptor survey also relate to the feasibility of executing the survey to yield information that is useful for the ERA. The proposed number of replicate visits (two or three per site) for the raptor survey is not sufficient to obtain a complete count of bird species and individuals. Most census methodologies (e.g., Hall, 1964; Robbins, 1981; Ralph et al., 1993) require four or more visits per site per season, on consecutive days.
3.12.3 Forest Bird Use Survey

Flaws in the proposed forest bird use survey (Appendix A.14-1) relate to the survey's objectives and inability to provide information regarding potential ecological risks posed by COPCs, as well as methodological deficiencies. Several suggestions are made below that should improve the likelihood that the survey will provide information that is useful to the ERA.

While the forest bird use survey, as proposed, could not be used to estimate ecological risks as a result of PCBs in the Housatonic River, it is not clear that this is, in fact, a goal of the survey. Indeed, the protocol indicates that the survey information may be used to help develop food web models and identify candidate receptor species and species of management concern. Consequently, the survey results may not be intended for use as a measurement endpoint. If that is the case, it is not clear that a survey of the magnitude described in Appendix A.14-1 is necessary for the ERA. If the proposed methodology were substantially modified, a forest bird survey could provide information that would be far more useful, site-specific, and ecologically relevant than could be provided by a generic literature-based TQ approach. Indeed, ChemRisk (1994) already surveyed forest birds within the study area and compared the observed density, diversity and composition to multiple reference areas monitored as part of the Breeding Bird Survey. Not only does that study provide a better methodological example of the use of forest bird survey methods in evaluating potential ecological risks, it also offers an existing site-specific data set that warrants consideration in EPA's ERA.

Several substantial modifications to the proposed methodology would improve the likelihood that the survey will provide information useful to the ERA. Such modifications include: concurrently surveying multiple reference areas; confirming the comparability of reference areas through habitat surveys; conducting four or more replicate visits at each site (two are proposed); and performing surveys on consecutive days during peak periods of territory establishment (i.e., the middle two weeks of May). The proposed timing of the visits (June and the first half of July) will greatly hamper the observers' ability to obtain an accurate count because observations of forest birds become much more difficult once the canopy is completely leafed out and because vocalizations decrease dramatically once territories are established. Furthermore, because the current survey design proposes to conduct the two visits during a six week period, it will be very difficult for the observers to identify breeding pairs or to replicate the area surveyed and the overall methodology.
Additionally, the breeding bird atlas of Massachusetts (Veit and Peterson, 1993) should be included among the references used to develop the initial list of birds potentially breeding at the study area. As previously discussed with respect to the raptor survey, the same observer should conduct each replicate survey, and that observer should be accompanied by a person designated to take notes. The methodology should include a training protocol that allows observers to practice the census method several days before the survey is actually conducted. Such practice runs help ensure that each survey visit shares a comparable quality of observation.

Finally, like many of the other protocols included in the work plan, no information is provided regarding methods of data analysis and interpretation. Those methods should be clearly defined prior to initiating the survey, because modifications to the survey design may be warranted to support data analysis and interpretation.

3.12.4 Marsh and Wading Bird Survey

The proposed marsh and wading bird survey (Appendix A.14-2) shares all of the limitations discussed above for the forest bird survey, including the survey’s objectives and inability to provide information regarding potential ecological risks posed by COPCs, and methodological deficiencies. In addition to the deficiencies associated with the number and timing of replicate visits, designation of a note-taker and a single observer for all replicate visits, reference areas, and methods to be employed for data analysis and interpretation, the work plan ignores the ready availability of data on the actual reproductive performance of great blue herons nesting within foraging distance of the Housatonic River. Productivity data (juveniles/nest) have been reported by the Massachusetts Division of Fisheries and Wildlife (MDFW, 1979; 1980; 1981; 1982; 1983; 1984; 1985; 1986a,b; 1987; 1989; 1991; 1996) and ChemRisk (1997) for these and other colonies throughout the state for the past 20 years. Such data will likely provide far more insight into potential effects of COPCs on this wading bird species than would be provided by TQs or the qualitative survey proposed in the Agencies’ work plan. The great blue heron data set includes results from numerous reference colonies (i.e., all colonies in Massachusetts that are beyond foraging distance of the Housatonic River), and has sufficient statistical power. Although there may also be limitations associated with this data set, associated uncertainties would presumably be taken into consideration as part of the weight of evidence evaluation conducted during risk characterization.
The work plan presents two appendices describing mammal surveys. The study of mammal use (Appendix A.15) focuses on small mammals, while the study of river otter, mink and bats (Appendix A.16) describes several different approaches to be employed to determine use by these mammals. The intended application of the mammal surveys is not clearly described in either appendix or within the text of the work plan. Consequently, it is unclear whether data will be collected entirely for descriptive purposes (for use in problem formulation), to select receptor groups, to help define study designs that will be used to quantify potential risks, or to describe potential ecological risks. In the latter case, evaluation of appropriate reference sites would be critical to interpreting survey results, but neither appendix discusses reference areas that may be evaluated. Based on discussions held May 6, 1999, it is our understanding that internal referencing will be employed at least for the small mammal tissue analyses; this approach should be noted in the work plan. Given the importance of prey, cover, and habitat in determining the presence of various mammal species, study areas must be closely matched. Toward that end, habitat surveys would be necessary to verify the comparability of the study areas. Additional concerns specific to each survey design are detailed in the following sections.

### 3.13.1 Study of Mammal Use

In addition to the limitations described above, the study of mammal use (Appendix A.15) is flawed with respect to the stated survey objectives and proposed methods of data evaluation and analysis. As discussed in Section 2.1, because this study has already been executed, weaknesses identified by GE (and others) cannot be remedied, unless the entire study is redone; consequently, if not redone, the study should be given little or no weight during risk characterization.

Appendix A.15 offers multiple contradictory objectives of the survey. While the stated objectives are to identify species of mammals that occur in the study area and to identify species in the study area of management concern (i.e., collect presence/absence data), trapped specimens will undergo chemical analyses and an attempt will be made to define a stressor-response curve. Because the study will only measure COPC concentrations in tissue and co-located soil samples, and will not evaluate actual responses associated with those stressor levels, it does not appear that development of a stressor-response curve will in fact be possible. Consequently, the only evaluation of responses (effects) will apparently be based on comparisons to literature-based no-effects levels.
As discussed during the May 6 meeting, since small mammal specimens will be collected, it would be relatively straightforward to dissect adult females to count embryos and placental scars, prior to submitting the samples for chemical analysis. This approach would allow evaluation of several site-specific measurement endpoints (i.e., sexual maturation, breeding, conception, and number of fetuses carried to term) related to reproduction in small mammals, adding a line of evidence that would be stronger than a literature-based TQ approach. Comparison of the number of placental scars per female across varying PCB tissue levels and relative to reference specimens would yield a stressor-response relationship by way of a site-specific line of evidence.

More information should be provided regarding the method of data evaluation and analysis to be employed. The protocol should identify the criteria for selecting MATCs, as well as the specific literature studies that will be used as a point of comparison, in order to ensure that the survey design, species, timing, age class, gender, organs, analytical methods, and congener composition are comparable. The technical basis for sampling 500 trap nights per site should be provided. The basis for the three selected small mammal trap sites should be described in greater detail than “based on guidance from U.S. EPA.” That is, the specific characteristics deemed most important to sampling site selection should be specified. If the small mammal trapping stations are within the Housatonic Valley Wildlife Management Area (HVWMA), then tissue samples should be analyzed for herbicides, since herbicide application is used as a resource management tool within HVWMA.

3.13.2 Study of River Otter, Mink, and Bats

The study of river otter, mink and bats (Appendix A.16) has problems related to the defined objectives and application of results in the ERA. Additional concerns relate to inadequate survey designs and insufficient detail.

It appears that the study of river otter, mink and bats will be conducted entirely for descriptive purposes and is not intended to generate estimates of ecological risk. However, the work plan provides no indication of how the information collected will be used in the ERA. If the observation of presence or absence of mink, otter and/or bats is solely to be used to select receptors of interest for subsequent TQ evaluation, that should be stated in the work plan. Because presence/absence data alone cannot be used as a measurement endpoint, the surveys may not prove useful to the ERA. Furthermore, the presence of mink and otter within the Housatonic River valley can be readily verified through historic state trapping records. Such detailed surveys are not necessary simply for the
selection of receptors of interest. We understand from discussions during the May 6, 1999 meeting that analyses of mink carcasses are no longer planned; the work plan should be modified to reflect this change.

Even given these limited and ambiguous objectives, the survey designs appear inadequate in several respects. First, Organ (1989) should be included among the references considered as part of the literature search, as he evaluated otter populations within the Housatonic River drainage and other drainages in Massachusetts. Second, the number of proposed replicate visits (two or three per site, depending on the survey design) does not appear sufficient to obtain a complete count of mammal species and individuals. Most census methodologies (e.g., Stickel, 1948; Drickamer and Paine, 1992) require four or more consecutive visits per site. Third, the technical and/or statistical basis for walking each mammal snow tracking transect two to three times should be provided.

Finally, the protocol lacks sufficient detail. For example, for the scent station surveys, the protocol should specify the survey dates. For the bat surveys, the protocol should provide information on the number of survey nights, survey dates, and survey duration. The specific sound analysis software should be named and tested, since its use will largely determine the success of the bat survey.

4.0 SUMMARY AND CONCLUSIONS

Of the multiple limitations of the work plan discussed above, five recur repeatedly throughout the proposed protocols and are most likely to interfere with the usability of the collected data for risk assessment purposes. These five major points are briefly reiterated below.

First, most of the assessment and measurement endpoints identified in the work plan focus on individuals, rather than subpopulations, populations, or communities (which are the proper focus of ERA). With few exceptions, survival, growth and reproduction of individuals are the assessment endpoints identified for the receptors. Additionally, existing data on populations and communities (e.g., great blue heron reproduction, passerine communities, fish communities) are not considered, and no rationale for excluding such studies is provided. The few field surveys proposed that do consider community-level endpoints (e.g., species richness) for amphibians, reptiles, and birds are not included among the selected assessment and measurement endpoints. Reliance on assessment endpoints that focus on individuals is not useful from a risk management standpoint, and will yield substantial overestimates of potential risks to higher levels of ecological organization.
Second, several proposed studies have no risk-related objectives. For example, several surveys appear to gather only descriptive data, rather than information that can be related to prediction of potential site-related risks. Other studies focus on tissue sampling and analysis, thereby treating bioaccumulation as an endpoint in itself, rather than ascertaining whether the populations are actually adversely affected.

Third, the work plan depends too much on comparisons to literature-derived benchmarks in place of site-specific information. Such benchmarks have numerous limitations. Literature-based benchmarks are based on conditions that differ from the Lower Housatonic River, with respect to exposure routes, COPC composition, species, and measurement endpoints. Consequently, estimates of risk based on such generic benchmarks require extrapolation and conservatism to account for the increased uncertainty. Additionally, such benchmarks rarely incorporate information from the entire stressor-response relationship, because they are usually limited to single point comparisons. Furthermore, most benchmarks reflect individual-level effects, rather than effects at higher scales of ecological organization. Finally, where site-specific field data exist or are proposed to be collected, generic benchmark comparisons are unwarranted and only dilute interpretation of the results.

Fourth, the majority of proposed field studies do not specify criteria used to select reference areas, the locations of candidate reference areas, or whether reference areas will even be evaluated. The proper definition of reference areas and a sufficient number of reference areas is critical to establishing causality for any observed differences in measurements. With few exceptions, the study designs do not explicitly name the reference areas that will be evaluated. In the rare cases where the work plan currently specifies reference areas, either the proposed reference area is inappropriate or sufficient justification is not provided to demonstrate that the reference area is appropriate. In cases where the work plan recognizes that reference areas will be evaluated, consideration of only a single reference area is generally proposed. Given that no single reference area will exactly match the study area with respect to all of the key criteria, evaluation of multiple reference areas will help minimize confounding factors in interpreting the study results.

Finally, the protocols presented in the appendices lack adequate detail, are poorly designed, and/or the studies themselves are unwarranted. Numerous flaws are common to the vast majority of protocols presented in Appendix A and B. First, the protocols are either ambiguous or do not discuss how results will be used and interpreted (e.g., will the population surveys be used for descriptive purposes
only or will they be compared to results from reference areas in order to evaluate risks). The sample sizes, timing and frequency of sampling events are either not specified or not justified. Analysis of lipids, wet weight, and dry weight concentrations are not planned in many protocols, although they are necessary for estimating bioaccumulation of PCBs and associated risks. Most appendices lack information on treatment of data, such as data reduction, presentation, statistical treatment, and application. The lack of detail provided prevents the complete evaluation of the feasibility, quality and scientific defensibility of many of the proposed studies. Consequently, there is the potential for substantial subjectivity in the interpretation of data generated. Furthermore, these gaps in the protocols may cause studies to be improperly executed and some critical data may not be collected.

5.0 REFERENCES


