N00129.AR.001044
NSB NEW LONDON
5090.3a

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
REGION 1
1 CONGRESS STREET, SUITE 1100
BOSTON, MASSACHUSETTS 02114-2023

May 13, 2004

Mark Evans, Remedial Project Manager
U.S. Department of the Navy
Naval Facilities Engineering Command
Northern Division
10 Industrial Highway
Code 1823, Mail Stop 82
Lester, PA 19113-2090

Re: Response to EPA Comments on the Draft Screening-Level Ecological Risk Assessment for Zones 4 and 7 of the Lower Subase

Dear Mr. Evans:

EPA reviewed the responses to EPA's comment letters dated February 27, 2004 and March 3, 2004 on the *Draft Screening-Level Ecological Risk Assessment for Zones 4 and 7 Lower Subase Naval Submarine Base - New London, Groton, CT*. The Navy responses are dated April 14, 2004. The numbering system used in the response to comments is retained herein. Detailed comments are provided in Attachment A.

Page 7 of RTC report states that various changes and clarifications have been made in the Workplan section for the DQO specifications on which the survey design estimates (sample sizes) are based. Some of those changes and corrections were also summarized in the RTC (4-14-04). Comments on those changes and corrections are summarized in Attachment B.

I recommend that the developers of the RTC report revisit the procedures proposed and described in Table 2 for the toxicity line of evidence. The specifics and the solutions for the question of interest for the project is not very clear to the developers of the statistical procedures to evaluate the toxicity line of evidence. The proposed procedures require more investigation and justification.

I look forward to working with you and the Connecticut Department of Environmental Protection to protect the environs of the Naval Submarine Base. Please do not hesitate to contact me at (617) 918-1385 should you have any questions.

Sincerely,

Kymberlee Keckler, Remedial Project Manager

Federal Facilities Superfund Section

Attachments

cc: Gareth Pearson, EPA, Las Vegas, NV
Mark Lewis, CTDEP, Hartford, CT
Melissa Griffin, NSBNL, Groton, CT
Bart Hoskins, USEPA, Boston, MA
Chau Vu, USEPA, Boston, MA
Jennifer Stump, Gannett Fleming, Harrisburg, PA

ATTACHMENT A

Number	Comment
2	The expansion of the uncertainty section proposed in the response will address EPA's comment as long as it discusses the spatial coverage and concentration gradients of fixed lab samples.
5	The response acknowledges that the TRVs for lead presented in Tables A-6 and A-7 are not correct, but states that the HQs are correct. The response indicates that the source for the TRVs used to derive the avian lead HQ is the EPA November 2003 Eco-SSL document. This does not appear to be correct. The EPA November 2003 Eco-SSL document describes the avian TRV derivation for lead as follows, "A geometric mean of the NOAEL values for reproduction and growth was calculated at 10.9 mg lead/kg bw/day. However, this value is higher than the lowest bounded LOAEL for reproduction, growth, or survival (Figure 5.1). Therefore, the TRV is equal to the highest bounded NOAEL, lower than the lowest bounded LOAEL for reproduction, growth, or survival, and is equal to 1.63 mg lead/kg bw/day. The NOAEL (1.6 mg/kg-d) presented in the response may be from the EPA soil screening guidance, but the LOAEL (16 mg/kg-d) value presented in the response is not derived in the EPA soil screening guidance. It appears as if the ORNL methodology of multiplying the NOAEL by 10 to derive the LOAEL was used to derive the LOAEL TRV. Table A-7 should be corrected to present NOAEL and LOAEL values with appropriate primary source citations. The hazard quotient should be revised as necessary.
6	The response indicates that the source for the TRVs used to derive the avian dieldrin HQ is the EPA November 2003 Eco-SSL document. The response does not appear to be correct. The EPA November 2003 Eco-SSL document describes the avian TRV derivation for dieldrin as follows, "A geometric mean of the NOAEL values for growth is calculated at 0.889 mg dieldrin /kg bw/day. However, this value is higher than the lowest bounded LOAEL for either reproduction, growth, or survival results. Therefore, the TRV is equal to the highest bounded NOAEL lower than the lowest bounded LOAEL for reproduction, growth, and survival results and is equal to 0.0709 mg dieldrin/kg bw/day." The NOAEL (0.0017mg/kg-d) and LOAEL (0.017 mg/kg-d) values presented in the response are not derived in the EPA soil screening guidance. Table A-7 should be corrected to present NOAEL and LOAEL values with appropriate primary source citations. The hazard quotient should be revised as necessary.

ATTACHMENT B

Table 1 *DQOs for Sediment Toxicity (page 7 of RTC):* How many samples will be collected from the 3 AOCs and a reference area? As requested earlier, please include formula to compute the p-value associated with the proposed use of the Dunnett's test.

Table 2 *DQOs for Invertebrate Toxicity Line of Evidence (pages 8 & 9):* As before, three decision rules (Survival decision rule, Growth decision rule, and Reproduction decision rule) have been proposed for invertebrate toxicity line of evidence. However, significant changes have been made in stating and testing of the hypotheses for the three (k=3) areas of concern (AOC) in the Lower Subbase Site. For example, in earlier reports (Report 1 and Report 2), it was proposed to compare the *proportions* of surviving amphipod *Leptocheirus plumulosus* at the end of 28-day bioassay test after exposure to sediments from Thames River Lower Subbase Site with the *proportion* of surviving amphipod *Leptocheirus plumulosus* at the end of 28-day bioassay in controlled laboratory experiment (as defined on page 13 of EPA 2001). All samples collected from an AOC (*e.g.*, Zones 4, 7, and outer Pier 1 of Lower Base) were supposed (*e.g.*, =6) to be included in the computation of proportion of surviving *Leptocheirus plumulosus* in the 28-day bioassay.

In the present RTC, survival proportion has been changed to mean survival per sample (i=1,2,3,4,5,6). The use of Dunnett test has been replaced by t-test to compare means of two populations (two single samples). It seems like 18 (=6x3) t-tests have been proposed for the 3 AOCs, one for each of the 6 samples (replicates) collected from these 3 AOCs. Thus, 18 t-tests will be performed for each of the 3 subordinate questions (54 tests in all) for the toxicity line of evidence. It will be difficult to draw a conclusion based upon the results of 54 (or even 18 for each question) t-tests.

It is not clear how will the proposed methods and tests for DQOs for toxicity line of evidence will work as discussed below. EPA is concerned that they actually may not work. The proposed procedures need further explanation and investigation.

Leptocheirus plumulosus **Survival Decision Rules (page 8)**

In the present RTC report, the proportions have been replaced by the means. For the Survival Decision Rule, the hypotheses have been stated in terms of mean number of *Leptocheirus plumulosus* surviving at the end of 28-day bioassay test after exposure to sediments from a **specific sample (i=1,2,3,4,5,6)** and the mean number of *Leptocheirus plumulosus* surviving the 28-day bioassay test exposed to the **laboratory controls (clarify how many controls will be used)**. It seems like that one test is being proposed for each of the six samples collected

(assuming 6 samples will be collected from each of the three AOC) from each of the three AOCs. Thus all together 3x6 = 18 sets of hypotheses and tests have been proposed for survival question of the toxicity Line of Evidence.

Questions and Concerns About the Procedures Described Above

It is not clear how the mean number of surviving *Leptocheirus plumulosus* for each of the 6 samples collected from the 3 AOC will be computed. How will the mean number of survivals based upon a single sample, i=1,2,3,4,5,6 be computed? How will this mean be computed based upon laboratory controls? **How many samples will be included in the laboratory controls (1 or more)?** Proper explanation of these hypotheses and tests, and the number of samples included in the computation of the mean number of survivals of *Leptocheirus plumulosus* should be included in the decision rule.

What happens when for a specific AOC (*e.g.*, Zone 7), 2 samples (out of 6 replicate samples) suggest that there are no differences (using a t-test?) in the mean survival of *Leptocheirus plumulosus* in lab controls and the sample under study (*e.g.*, i=1,2), and the other 4 samples (*e.g.*, i=3,4,5,6) suggest that the mean survival of *Leptocheirus plumulosus* for samples from AOC is smaller (using a t-test as proposed in RTC?) than the survival mean of *Leptocheirus plumulosus* in lab controls?

How will one make a decision about the mean number of survivors of *Leptocheirus plumulosus* for that AOC under study?

Which are the statistical populations under study? How many populations do we have here (3 or 18)? All these questions need to be answered for each of the decision rules for toxicity line of evidence as described in Table 2 (pages 8-9 of RTC report). Typically, for each AOC, the mean number of survivals should be computed based upon the total number of samples (=6 here) collected from that AOC. Is the sample size 6 based upon the corrected formula as mentioned on page 16 of the RTC?

***Leptocheirus plumulosus* Growth Decision Rules (page 8)**

Just as for the Survival Decision Rule, it is proposed to compare the mean growth of *Leptocheirus plumulosus* after exposure to sediment from a sample (i=1,2,3,4,5,6) from an AOC with the mean growth of *Leptocheirus plumulosus* after exposure to laboratory controls (how many)? These tests will be performed separately for each of the three AOCs. It seems like 18 t-tests have been proposed.

Please explain how many samples will be included in the laboratory controls? Proper explanation of the number of samples (from an AOC and also the number of lab control samples) included in the computation of the mean growth should be included in the decision rule.

Which are the statistical populations under study? How many populations do we have here (3 or 18)?

How do we compute of the **mean number** of survivors and the **mean growth** per sample (i=1,2,3,4,5,6) in 28-d bioassay? Please explain.

What happens when for a specific AOC, 3 samples (*e.g.*, out of 6 samples) suggest that there are no differences in the mean growth of *Leptocheirus plumulosus* in lab controls and the sample under study (*e.g.*, i=1,2,3), and the other 3 samples (*e.g.*, i=4,5,6) suggest that the mean growth of *Leptocheirus plumulosus* in samples from AOC is smaller than the mean growth of *Leptocheirus plumulosus* in lab controls?. **How will one make a decision about the mean growth of *Leptocheirus plumulosus* for that AOC under study?**

Also, the sample size formula (2) given on page 5 of the earlier Navy's Draft Report 2, is to compute the minimum sample size needed (to achieve the DQOs, α, β, Δ *etc.*) to compare the means of two populations (an AOC vs. control, reference). How does this sample size (=6) fit in the present application/setting?

***Leptocheirus plumulosus* Reproduction Decision Rules (pages 8-9 of RTC)**

Just as for the Survival Decision Rule, it is proposed to compare the mean number of offsprings produced *by Leptocheirus plumulosus* after exposure to sediment from a sample (i=1,2,3,4,5,6) from an AOC with the mean number of offsprings of *Leptocheirus plumulosus* after exposure to laboratory controls? These tests (two sample t-tests have been proposed) will be performed for each of the three AOCs. As before, 18 t-tests have been proposed.

Please explain how many samples will be included in the laboratory controls.

Which are the statistical populations under study? How many populations do we have here (3 or 18)?

How do you compute of the **mean number** offsprings per sample (i=1,2,3,4,5,6) in 28-d bioassay?

What happens when for a specific AOC, 3 samples (*e.g.*, out of 6 samples) suggest that there are no differences in the mean reproduction of *Leptocheirus plumulosus* in lab controls and the samples under study (*e.g.*, i=1,2,3), and the other 3 samples (*e.g.*, i=4,5,6) suggest that the mean reproduction of *Leptocheirus plumulosus* in samples from AOC is smaller than the mean reproduction of *Leptocheirus plumulosus* in lab controls. **How will one make a decision about the mean reproduction rate of *Leptocheirus plumulosus* for that AOC under study?**

Also, the sample size formula (2) given on page (5) of the earlier Navy's Draft Report 2, is to compute the minimum sample size needed (to achieve the DQOs) to compare the means of two

populations using a t-test. How does this sample size (=6) fit in the present application and setting of comparing means using single samples? Will the statistics (*e.g.*, t-test *etc.*) be computed based upon a single sample instead of 6 samples?

Sample Sizes in Design Document as Related to Toxicity Line of Evidence (page 16)

Use of t-tests based upon single samples is re-iterated on page 16 of RTC report. It is stated that, the question of interest for this project suggests testing each sample (perhaps a total of 6 samples from each of the 3 AOC) at a given site (3 AOC sites) against a laboratory control. Two sample t-tests are used in this situation to compare two sample means (based upon two single samples - one from a specific AOC and other(s) represents a lab control). Hence replicate calculations (sample size?) are based upon t-tests.

If the number of replicates for t-test obtained using corrected equation (2) and (3) comes out to be six, according to the above statements, for 6 samples from an AOC, will 6 t-tests will be performed, one for each sample?

The sample size equation as given in equation (2) of Report 2 applies for 2 population comparisons. Six (if the sample size equation (2) yields a number 6) samples need to be collected from one population and 6 samples need to be collected from the other population (*e.g.*, lab control). Using the data thus obtained, depending upon the underlying assumptions, an appropriate single t-test can be performed to compare the two means of the two populations - AOC and Control.

But in the present context, the Navy is proposing to perform 6 different t-tests (one for each of the 6 samples from an AOC) to compare the various hypotheses for toxicity line of evidence.

EPA has several questions about the adequacy of such a test as stated above in the three DQO sections. It is not clear how many lab controls will be used to evaluate the toxicity line of evidence. Using equation (2), the mandate is to use 6 (if equation 2 yields n=6) lab control samples. How will these 6 samples be used in the proposed hypothesis testing based upon single sample t-tests, one for each replicate sample?

There is no theoretical connection between the number of replicates (*e.g.*, =6) and the proposed use of the t-tests for toxicity line of evidence.

Furthermore, as stated earlier, how will a decision be made about the various subordinate questions for this line of evidence? When for a specific AOC, 3 samples (*e.g.*, out of 6 replicate samples) suggest that there are no differences in the mean reproduction (survival or growth) of *Leptocheirus plumulosus* in lab controls and the sample under study (*e.g.*, i=1,2,3), and the other 3 samples (*e.g.*, i=4,5,6) suggest that the mean reproduction (survival or growth) of *Leptocheirus plumulosus* in samples from AOC is less than the mean reproduction (survival or growth) of *Leptocheirus plumulosus* in lab controls, how will one make a decision about the mean

reproduction (survival or growth) rate of *Leptocheirus plumulosus* for that AOC under study? It is important to rethink and investigate these issues further. Perhaps some other more appropriate test can be used to statistically evaluate the toxicity line of evidence.

Table 3 (page 9) DQOs for Food-Chain Modeling Line of Evidence

How many samples will be collected from the three AOCs? How many samples will be included from the reference area? Are these given in the revised design document?