N00296.002888
MOFFETT FIELD
SSIC NO. 5090.3

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
REGION IX
75 Hawthorne Street
San Francisco, CA 94105-3901

July 26, 1996

Mr. Stephen Chao
Naval Facilities Engineering Command
Engineering Field Activity, West
900 Commodore Way, Bldg. 101
San Bruno, CA. 94066-2402

Re: *Draft Final Phase II Site-Wide Ecological Assessment Report*, dated May 1996

Dear Mr. Chao,

The U.S. Environmental Protection Agency (EPA) has received the subject document and the associated response to comments. We appreciate the quantity of work completed since the submission of the draft version of this document. It shows the results of many meetings and conference calls. EPA believes that reasonable decisions regarding questions of remediation can now be made. Most of the summary figures, e.g. the Thiessen polygons, presented in the report are especially helpful in providing the necessary information.

In general, bioassays provide a more accurate risk estimate than hazard quotients. There is far too much uncertainty in hazard quotient calculations. If one wanted to reach the most accurate conclusions about risk in these ecological areas, then more bioassay samples would be required. We understand through conversations with you that certain ecological areas have been targeted for remediation in the Station-Wide Feasibility Study (SWFS). While questions about this assessment still remain, we believe that the weight of evidence approach adopted clearly supports remediation in many of the ecological areas. Our attached comments are directed towards pointing out specific problems that, if larger budgets and time permitted, would be desirable to pursue. How much remediation to perform may still remain a question. But we agree that resources should be put towards remediation. If the Navy decides not to remediate areas that show high risk towards any receptors, we may end up revisiting the results in this ecological assessment. But at present, we believe the ecological assessment at Moffett is sufficient to move towards remediation. Responses to the attached comments should be provided for clarification, but with remediation to occur, please do not delay the submittal of the SWFS. If you have any questions, please call me at 415-744-2385 or Clarence Callahan for technical direction at 415-744-2314.

Sincerely,

Michael D. Gill

Michael D. Gill
Remedial Project Manager
Federal Facilities Cleanup Office

cc: J. Chou (DTSC), K. Eichstaedt (URS), S. Olliges (NASA), M. Rochette (RWQCB), P. Strauss (MHB), K. Walsh (MW) (email), M. Young (PRC) (email)

Printed on Recycled Paper

AR
2888

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
REGION 9
75 Hawthorne Street
San Francisco CA 94105-3901
July 25, 1996

MEMORANDUM

SUBJECT: Moffett Federal Field Review

FROM: Clarence A. Callahan, PhD
Biologist, Technical Support Section (H-9-3)

TO: Michael Gill, Remedial Project Manager
Navy Section (H-9-2)

I have read the document and provide the following comments. This material is provided to you with the understanding that I have not had a chance to talk to Laura Valoppi and Susan Gladstone about the document or my comments. I am planning to discuss these comments with them and I may revise these, not drastically, but to make them clearer and more usable.

Overall risk statements.

Range of Risk - It is clear from the figures in chapters 9 and 12 that Moffett Field is characterized as having a range of areas from low risk to very high risk, at least by the methods presented in this document. If these proposed risk estimates withstand a validation phase, the challenge will be to determine the cleanup level for the contamination and move on to the determination of exactly what areas need to be remediated.

Closure of the process - The document does not complete the process. The various figures with accompanying text should be integrated and presented in an overall risk figure and text for each of the areas, the Eastern Diked Marsh, the Storm Water Retention Ponds, the Ditches and Channels and Upland Areas. Chapters 9, 10 and 11 emphasize the further development of the hazard quotient, rather than using the Phase II hazard quotient results (i.e., predictions of effects by the HQ approach) to identify the areas of the effort for confirmation and to reduce the areas with high uncertainty. The material presented in the risk estimation of the risk characterization phase should be the "results" or a "presentation" of the data that forms the basis of the effort. However, after the Phase I effort, the Navy completed few bioassays and added few additional samples to validate the Phase I results, leaving the process with higher levels of uncertainty than expected.

The risk description effort of the risk characterization is the "discussion" or the explanation of the effort based on the data gathered for the overall risk assessment. The Navy added a structure to address the "weight of evidence" approach for the discussion of risk description which is integrated into the last chapters.

Organization - The organization as presented is Chapter 9-Risk Estimation for all areas; Chapter 10-Risk Description for surface water; Chapter 11-Risk Description for sediments; and Chapter 12-Risk Description for soils. The two components of risk characterization are risk estimation and description, risk summary is part of the risk description (EPA, 1992).

Chapter 8

CHARACTERIZATION OF POTENTIAL ADVERSE EFFECTS ON ENDPOINT RECEPTORS

Estimation of risks - The approach used by the Navy for this task was the hazard quotient (HQ) based on "...the surface water chemical data..." The Navy then, "...reviewed the bioassay results to determine if there is a relationship between the bioassay responses and the COPECs found in the bioassay media that may drive any risk. In addition, the results of the bioassays are compared to the HQ and HI data as part of a weight-of-evidence evaluation." The stated purpose, "...was to determine if a relationship exists between the two data sets that will help to identify potential risks..." "Bioassay data are used to identify NOAEC levels for specific areas, and these levels are compared to the biological responses and chemical constituents in the other bioassays from the same areas."

The use of the HQ as a "definitive tool" is faulty at best, is limited in technical soundness, and cannot be supported by the data to estimate risk at a definitive level. The estimation of risk is best demonstrated by bioassays and direct measurements of contaminant levels from the site, as opposed to the use of HQs. The HQ approach is very useful for screening of COPECs early in the process, but has very limited utility in the later stages of the risk assessment process because of its use of data that are of limited quality i.e., high uncertainty and this is demonstrated in the information provided by the Navy in this document. There has been a strong opinion from the Agencies for the use of bioassays in this process, particularly at the Phase II level. Early EPA (1986) documents state the limitations of the HQ process, the DTSC (1994) risk assessment guidance document uses the HQ process as a screening tool but not for the latter stage of the process.

The Navy presents an argument for the justification and interpretation of the use of the HQ that is contradictory to the acceptable use and interpretation in Region 9. For instance, p8-3, "If the HQ or HI is one or greater, then the interpretation is that adverse effects are possible. This is compatible with the use and interpretation by EPA in Region 9, however, the statement by the Navy, "However, there is an implied increase in the potential for harm with an increase in the HQ is not compatible with Region 9 use and interpretation". Barnthouse et al (1986) as cited in the Navy text on p8-3 provides the best recognized use of the HQ, "...a valuable **screening** (emphasis added) technique..." And the stated disadvantage, "...the differing sensitivities for various species and endpoints, and the need to extrapolate toxicity data and predictions among endpoints, taxonomic groups, and from laboratory to field situations (EPA 1991)" is the primary reason for any use other than screening.

p8-1, The use of the hazard quotient introduces a number of problems including the large amount of uncertainty about the "relative value" of the ratio. The quality or value of this quotient as an interpretative tool, of course, depends upon the quality of the input data. Traditionally, the US EPA views the use of this ratio with caution, especially in Region 9 where it is accepted in a screening approach, but find its value highly over rated as a tool for definitive actions in the latter stages of any ecological risk assessment. The Navy is advised to review the limitations of the HQ approach in the Office of Toxic Substances document by Urban et al (1986).

Bioassays were used in the Phase II effort, however the number was limited for programmatic reasons by the Navy with possible insufficiency of data for all areas of the potentially impacted areas. The experimental design was lacking in that instead of developing the exposure-response throughout the entire "risk" gradient, the Navy tested a selected few samples. By examining "random" sample results, there is little ability to achieve a correlation of the response with the contaminant risk gradient either for single contaminants or for multiple contaminants. The bioassay data are not presented as a group showing the range of responses with the range of concentrations so that site specific results can be shown. To the Navy's credit, however, this was completed for the Northern channel.
p8-4, There is little reason to be concerned about the inability of "single species" tests not being representative of measurement of ecological effects to higher level structures, such as populations, communities, and ecosystems when the single species tests are sufficient to make decisions about impact when performed in an adequate experimental design. Single species performed correctly are much more useful compared, for instance, to an assessment of community structure analysis that does not yield any definitive answers. I disagree with the Navy that a multispecies approach was taken at MFA. The use of a number of single species tests each with a single receptor does not constitute a multispecies approach in the sense of population and certainly not a community.

p8-5, **Hazard Quotients**. "Reasonable maximum exposure" (RME) is not a term used for ecological risk assessments, however it may be appropriate for human health risk assessments. The term should be eliminated from the text pertaining to ecological risk assessment.

Potts, 1995 is not an employee of Region 9 and does not set risk assessment policy for Region 9. "Total" rather than dissolved concentrations should be used to estimate risks to aquatic receptors. Although the dissolved fraction may be the most "bioavailable fraction", water column species are impacted by contaminants through the ingestion pathway in addition to other pathways, therefore all of the component that represent the "total ingestion" including that attributed to contaminants adhered to suspended particles must be included.

p8-7, **Stressor-response analysis**. Of the four objectives of this section that help to "identify evidence of causality" identifying the COPECs that have the potential to adversely affect benthic receptors and to identify the location of the adverse ecological effects have the most utility, whereas any efforts to evaluate the "magnitude" of the adverse ecological effects and to evaluate the "duration" of the adverse ecological effects are less likely to directly address the overall risk process.

p8-9, I don't see any value in the tabular entry for the "Proportion HQs Greater Than 1."

p8-10, **Bulk sediment hazard quotients**. Bulk sediment hazard quotients with HQs above 1 indicate a potential problem; there is little evidence that the severity of risk is indicated at any higher or lower level when HQs are observed above 1. The "benchmark" is 1.0 and the effort should be to insure that the estimate of the HQ is as certain as possible. A critical deficiency in this information and the resultant estimates of risk is related to the determination of the "Exposure Point Concentration" as listed for Table 8-4. The use of the 95% UCL seems appropriate for those data with samples above 5, however, for those data with samples less than 5, the maximum value should be used instead of the mean. Calculating the mean including all of the data with non-detects lowers the mean value in a biased manner. Although, the ranking of HQs may be a logical attempt to achieve some kind of summary process for these data, I'm

not sure how far the process can be carried forward i.e., risk drivers.

p8-10, **Pore water hazard quotients**. I'm a little confused, from the text on this page. Only two samples were analyzed for inorganics and organics, PWNC-18 and ELWL-22, whereas Table 8-5 indicates that there were "9" analyses. How were these samples distributed? Were the samples from each location split for analysis according to some design e.g., dilution series?

Western Diked Marsh

p8-11, **Bulk sediment hazard quotients**.

Although, the ranking of HQs may be a logical attempt to achieve some kind of summary process for these data, I'm not sure how far the process can be carried forward i.e., risk drivers.

Pore water hazard quotients. From Section 6.5.2.1, Theoretical pore water HQs were based on estimates using the EPA (1993, 1994) equilibrium partitioning model. The "organic" portion of this approach seems to have misused the model in "...assumes that the EqP theory explains behavior of **all organic COPECs**" when at best, the model is aimed at "non-ionic" organics. What effect does this have on the prediction of pore water concentrations of ionic organic compounds?

Apparently, the pore water concentration for each is highly dependent upon the COPEC distribution coefficient (K_d) with parameters such as pH, iron oxide concentration and organic matter having a large influence on the final estimate of the pore water contaminant concentration, here called the "theoretical pore water concentration." The Navy qualifies the estimates with the statement, "The accuracy of the theoretical PW COPEC concentrations is the largest uncertainty associated with the appropriateness of the PW COPEC exposure point concentrations. The Kd values...do not reflect site-specific conditions. Nor do the Kd values reflect other important parameters, such as clay content, oxidation-reduction conditions,and major ion chemistry."

Were the predicted values compared to measured values for MFA and where are these data presented? Even with the statement, "The limited PW COPEC data collected under the MFA Data Gap Investigation showed COPEC concentrations less than predicted values" the fundamental data needed to "scope" the sample specific results were not used for each sample, thus the one sample presented as evidence of "overestimation" of pore water concentrations is of limited value in predicting these data for all samples.

Ditches and Channels

p8-13, **Bulk sediment hazard quotients**. Several hazard quotients exceed the quotient of 1.0.

p8-14, **Pore water hazard quotients**. The same precautions about the development of theoretical values for pore water concentrations for other sections apply here as well.

p8-14, **Summary of bulk sediment and pore water hazard quotients**.

The summary data in this section and the accompanying tables identify several organics and

metals as "most likely" contaminants of concern, moving them from the "potential" to the most likely level. In fact the Phase 2 results should be the identification of contaminants that are no longer "potential" but are identified as contaminants of concern. This is never done in this document.

p8-15, The sediment and pore water bioassays and chemistry should be used to determine/define the relationship between "how much exposure (concentration) produces how much effect (response)". If this is suggested, it is not clearly presented. This was the stated goal of the bioassay work from the first suggestion that they need to be performed. "The bioassay results indicate that MFA sediments adversely affected some bioassay organisms...The chemistry data showed the presence of several toxic COPECs in bulk sediment and sediment pore water." Although "Statistical correlation of bulk sediment toxicity bioassay response data with bulk sediment COPEC concentration data provided limited information on the COPECs that may explain adverse responses observed in the bioassays", the most important information gained by these tests is that **the contaminants are bioavailable i.e., the sediment samples caused a positive response in the tests employed for this evaluation**. I agree that if a next step is taken, they should include tests to identify the specific contaminants that cause the responses in the bioassays, perhaps "toxicity identification evaluation."

p8-17, **Synoptic bulk sediment bioassay data and bulk sediment chemistry data**. The amphipod showed statistically significant reductions in survival for all samples except one. In these samples, survival was reduced from 18 to 56 percent, as compared with the control. The polychaete showed no statistically significant reductions in survival for any samples. One sample showed a 30 percent reduction in growth that was statistically significantly less than the control. It is these data that should be combined and examined in a scatter diagram to evaluate for trends and relationships between exposure and effect. Table 8-14 shows that among the amphipod results, sample SSRP-31 should be considered a good candidate for a "field reference" location because of the low level of mortality, very low toxicity (93% survival) and high burial, very high suitability of sediment texture (100% reburial).

The polychaete growth should be expressed as a "percent of control" rather than mean weight because as shown, there is no point of reference; a benchmark like the individual test control growth is needed.

p8-17, **Synoptic pore water bioassay data chemistry and pore water chemistry data.** For the two samples where pore water chemistry and bioassays were performed, SSNC-18, echinoderm survival was impaired at pore water dilutions of 100 to 68 percent and echinoderm development was impaired at pore water dilutions of 100 to 34 percent. For SSNC-19, echinoderm survival and development were impaired at pore water concentrations ranging from 100 to 25 percent for survival, and from 100 to 12.5 percent for development. The reference toxicant, copper, used in these bioassays showed that at 9ug/L, sand dollar fertilization was reduced by about 50 percent, and at 18 ug/L copper, no fertilization occurred. As noted above, the concentration of copper in SSNC-18 and SSNC-19 was 60 and 100 ug/L respectively.

Table 8-15 does not present the appropriate information. If the intent is to compare the pore water concentrations with the HQs, then the detection limits should be provided for those contaminants shown as "ND" to get some idea of the levels of contaminants. Also, to further summarize and possibly infer some meaning from these data, why not present the two columns of data in a scatter diagram, COPEC (*ug*/L) X COPEC HQ? A scatter diagram would be much

more instructive than a mere listing of these data.

p8-18, **Pore water sulfide and ammonia data**. The citation, "EPA, 1987" is not listed in the reference section. The criterion of 2 ug/L for H_2S cannot be verified, however, SSNC-18 had 24 mg/L of sulfide and SSNC-19 had 91ug/L. Ammonia also exceeded the EPA water quality criteria (of 6.9 mg/L, acute and 1.33 mg/L, chronic for a 4 day average) at both of these sites. A bioassay or an ammonia "control" would have been very instructive to evaluate the potential effects of ammonia for these samples.

p8-19, **Bulk sediment AVS and SEM data**. It is not clear what is actually measured at these sites because, although, the importance of AVS and SEM measurements is discussed by the Navy, it appears that SEM and "bulk sediment metals" are presented in Table 8-16. The statement, "The ratio provides a general measure of the proportion of each metal that may be sequestered by AVS" is not clearly obvious and may not be relevant. There is some doubt about the collection and handling of the samples to obtain AVS and SEM data as indicated by the statement on p8-20, "...concentrations of chemical parameters between subsamples may be substantially different because of within-sample heterogeneity - even after mixing -..." These sediment samples should not be mixed when collected for AVS and SEM measurements. The description of sampling techniques states that water in the sample container was replaced by sediment to avoid mixing, which is the correct information. Further, on p8-21,"...the AVS concentration of a metal occasionally exceeds the total concentration of a metal." which does not suggest that proper handling techniques were practiced, thus, possibly voiding these data.

p8-21, last par., Regarding the statement, "The ratio of total SEM-to-AVS provides some evidence that these metals might be found (should be available) in the sediment pore water". Table 8-17 shows that five samples, SSRP-27, -28, -30, -32, and SSWL-22 have metals that are available as indicated by the ratio greater than 1.0. What is the relationship of these data to other measurements of metals for the same locations? Do these data compare well or are there large differences with measured values in other measurements?

AVIAN RECEPTORS

p8-24, Last par, The HQ_4 estimate is based on the maximum detected concentration at the site **and** as a conservative estimate of risk, this concentration should be considered over the entire habitat and integrated with the nearest neighbor sample results as was done for the presentation of the Thiessen polygons.

p8-26, **Relationship of measurement endpoints and assessment endpoints**. The written material in this paragraph does not convey the generally accepted meaning for assessment endpoints, measurement endpoints and the data needs to support their evaluation. The Table 3-1 is, however, a good description of these relationships. The HQ is not a measurement endpoint, however as suggested, it is representative of measurement endpoints for acute toxicity (mortality) and growth (change in weight). These measurement endpoints must be supported by the collection of data e.g., through bioassays, for mortality and growth over time and the HQ is a convenient manner to express the relationship between the observed data and the benchmark that is reported to result in a significant change in the measurement endpoint. If the data collected, when compared to the benchmark, produces a quotient above 1.0, then the assessment endpoint is expected to be impacted. The key word here is "expected" unless the bioassay was performed with site specific samples, the "numbers" are generally best estimates for the process and caution

is raised until the results can be verified.

The results of this section are best in tabular output that shows the best estimates for the exposure data compared to the benchmarks such as that in Tables 8-18 through 8-24.

CHAPTER 9

CHARACTERIZATION OF ECOLOGICAL RISK

The "Framework" document states that, "Risk characterization contains two major steps: risk estimation and risk description". "Risk summary" as listed on p9-1 is not the same as risk description.

p9-1, Purpose - "It is the objective of this section to characterize the potential adverse effects that may occur because of exposure to MFA-related constituents and to identify, where possible, those constituents that are the primary cause of that risk." The key phrases in the statement of purpose are: 1) "**to characterize the potential effects**" and 2) "**to identify...those constituents that are the primary cause**" both of which are needed to fully characterize the risks in the ecological risk assessment process.

The characterization of risks in the ecological risk assessment process involves two major efforts: 1) ESTIMATION OF RISKS AND 2) DESCRIPTION OF RISKS. Very generally, the estimation of risks involves: 1) the identification of the contaminants on the site; 2) the receptors on the site; and 3) the intersection of the contaminant concentrations that result in a significant impact to the receptors. Specifically, the process involves the collection of site specific data to estimate and describe the ecological risk at Moffett Field with the contaminants and receptors that are specific to Moffett Field.

The Navy's effort for risk estimation (from p9-1) involved the comparison of single effect and exposure values and comparison of distributions of effects and exposure. The Navy presented the results of risk estimation for each of the major receptor groups, surface water and sediment receptors using the hazard quotient approach with some bioassay results.

p9-2, The suggested framework (pp9-2 and 9-3) for evaluating the quotient results as cited from Menzie et al, 1993, is not an approach that is practiced in Region 9. Also, the statement in the second paragraph, "...with significant input from the EPA Region I Biological Technical Assistance Group (BTAG)..." is taken out of context and may lead some readers to a false assumption that the "framework" referenced above has the approval of the Region I BTAG when that is not the situation. From the cited paper, page 90, a lead off sentence in the section titled, "Selection of ecological receptors" the sentence reads, "The EPA Region I BTAG group (sic) had substantial input in this process." This input and process refers to the site surveys for the various case studies listed in the particular section of Menzie's paper, rather than the approach suggested for "interpretation" of the hazard quotient. The "toxic quotient" (Menzie et al, 1993) is discussed in the section on "Risk Characterization" (p94) of the Menzie paper and emphasizes the use of this approach at two sites or case studies listed in the paper which is hardly an endorsement of the approach for wide application, thus its use at Moffett is not warranted.

The acceptable Region 9 approach is to consider any hazard quotient (toxicity quotient) above

1 as a candidate for verification of the input data for its estimation. The basis of course, is that these data are the best that are available, otherwise they will have to be verified to reduce the uncertainty.

Eastern Diked Marsh.

p9-3, Contaminants with HQs above 1.0 are shown in Table 8-1 as Total Aroclor, chromium, copper, lead, manganese, nickel, silver and zinc. The sample locations include SWWL-22, SWWL-23, SWWL-24, SWWL-25, SWWL-26 and SWWL-003. The summed HQs for this habitat is 150.1.

p9-3, Eastern Diked Marsh. The Reasonable Maximum Exposure HQ is a term that is not used in ecological risk assessment and has no meaning, it should be eliminated from this discussion. The HQ based on the 95% UCL of the observed data is appropriate, however what is the logic for calculating this value with only one detected value, e.g., selenium and silver, the maximum value should be used? What does a "--" indicate?

It appears that all of the surface water samples from the Eastern Diked Marsh produce a hazard index above 1 using the methods and data as presented by the Navy. There is little reason to expect a fish, an insect, or an amphibian to have a higher level of protection by using the dissolved fraction rather than the total contaminant concentrations. Because of the approach (HQ) utilized by the Navy at this time in the risk assessment process, the total contaminant concentration is the most conservative.

The reference to the dissolved fraction being the more appropriate form to use when evaluating potential impacts (for water column species without direct ingestion) should be evaluated with the data on hand or this remains a high uncertainty. This is the kind of uncertainty that can easily be verified during the effort to estimate or calculate impact and risk. The dissolved fraction should be used for interpreting the potential impact to algae and the total should be used for interpreting the potential impact to fish and invertebrates.

The last sentence on p9-4 in the first paragraph is misleading and very general without reference to any particular receptor and endpoint, it is virtually meaningless ["The COPECs associated with the suspended material are expected to be significantly less available than the dissolved form (EPA 1995)"].

Storm-water Retention Pond.

p9-4, There were three COPECs copper, manganese, and nickel with HQs above 1.0 and lead with an HQ of 0.95. Lead would be expected because of the flow patterns for surface water and its high level of concern in the "upstream" area, the Eastern Diked Marsh. The distribution of HIs above 1.0 is wide spread throughout the ponds (SWRP-2, SWRP-4, SWRP- 9, SWRP-19, SWRP-27, SWRP-28, SWRP-29, SWRP-34 and SWRP-35) for a combined HI of 13.

Ditches and Channels.

p9-4, It appears that the predominant contaminant class is metals and a limited number of these are throughout all the areas. Twelve of 28 locations have HQs above 1.0. How was an HI estimated at 0.0? There were three COPECs copper, manganese, and zinc with HQs above 1.0

and lead and nickel with an HQ of 0.946 and 0.975 respectively. The summed HQs expressed as an HI is 29. Although 16 locations were reported as having an HQ less than 1, 12 had HQs above 1.0: SWMR-001, SWMR-002, SWMR-003, SWNC-001, SWNC-002, SWNC-003, SWNC-18, SWND-001, SWND-002, SWND-004, SWND-005 and SWND-006.

Interpretation of bioassay responses response data.

P9-5, Interpretation of Bioassay Response Data. The data from the surface water samples that the algal, mysid and Menidia tests are not obvious if presented. These data are needed to compare to the test results. Based on the results observed there were two samples that exhibited a response in bioassays that might be considered a potential impact to the test species. The location, SWRP-32, indicated some toxicity for the mysid and the Menidia, although not clearly toxic and the sample location SWNC-18 showed some toxicity with mysid alone.

p9-5, The objective of this Navy effort was "...to determine if they provide an insight to the potential risks to the surface water receptors at MFA." Even though the algal bioassays exhibited no statistical reductions in growth in any of the surface water samples, the contaminant concentrations in these tests should be graphically compared to the responses to show the relationships between water concentration and response of the algae. The same is true for the mysid and the Menidia test results. The results observed for the mysid results are not uncommon for situations where the concentration of exposure is near the low end of the response. Perhaps the test should have been repeated to clarify the results because no explanation is offered for the anomaly.

The Menidia results present another opportunity to determine the concentration relationship to the response. What was the actual concentration of exposure in the two levels that produced the 31 and the 25 percent reductions in growth? These concentrations along with the others should be used to develop the response gradient to the exposure concentrations and in turn be used to interpret the other observed concentrations for surface water.

The loose use of "sensitivity" as on p9-6, should be avoided. Although there were no responses (reported as significant) there were responses that are applicable to the questions at hand. With the "limited" contaminants in the surface water, perhaps there should not be an expected response other than the ones observed, if the correct test species were used. To describe something as being "low sensitivity" in this case, does not take into consideration the purpose for using these test organisms. If the tests represent the receptors and endpoints that are appropriate for this site, and the tests were performed within the design, then the results should indicate the data that can be used to interpret the assessment endpoints for this process whether or not there is a "significant" effect.

Correlation of bioassay responses with COPEC concentration.

pp9-6 and 7, The comparison of the bioassay data to "other samples" as mentioned on this page and shown in Table 9-2 is inadequate and provides very little if any information. The raw data should be plotted for each test and then combined in a single graph to evaluate the possibility of a trend being exhibited for concentration and response. The single test results do not provide sufficient data to evaluate this relationship. The only relevant result discussed is the potential for arsenic to be a "risk driver" but the discussion does not compare these results to any literature values, which is a shortcoming.

p9-6, **Correlation of Bioassay Responses with COPEC Concentration**. Does Table 9-2 reflect the actual chemical analysis of the same sample water i.e., split sample that the test was performed on or are these data summarized from the separate water data? The bioassay data represent laboratory (in this case) results from samples collected from the site and exposing test organisms that are best representative of the site situation being tested, whereas the hazard quotient is the ratio of the 95% UCL estimated from sample data throughout the site and the dose that produces a significant effect generally the no observed effects level. The bioassay represents direct results from the sample collected at the site and provides an estimate of the effects of the biologically available contaminants in that sample from the site. The hazard quotient is not a direct method, only uses the contaminant concentration measured at the site and does not provide any information of the biologically availability of the contaminant. In fact, much of the time, the dose is derived by estimating the uptake into the surrogate receptor using literature derived transfer factors which are most often modified by relationships for characteristics such as total organic matter, pH and water hardness. Obviously, the use of bioassay data provide more direct information and are superior to "modeling" a dose for the hazard quotient approach.

Where in the document were the HQs compared to the bioassays? If the Navy is suggesting or has evaluated the relationship between the HQs and bioassays as suggested, the comparisons are not clearly provided. I doubt whether or not such a comparison is useful because of the data input to the HQs. The comparison as suggested and presented in Table 9-3 is a continuous variable (HQ) predicting a discontinuous variable (yes or no answer). For instance, the Navy suggests that the HQs were considered as the predictor (independent variable) and the bioassays are the predicted variable (dependent variable). Table 9-3 only shows the bioassays that were significant which is very limiting and not logical.

Correlation of Bioassay Responses and COPEC Hazard Quotients.

p9-7, The attempt to compare a continuous variable, HQ with a yes or no answer, to bioassay results as shown is not technically justifiable and is incorrect as presented. Bioassays are superior to the HQ for determining the potential for impact, in fact, it is the most direct method for determining the relationship between contaminant concentration and receptor response. There are no data presented in this section to suggest that HQ results can be used to predict the bioassay results and it is apparent that the Navy reached the same conclusion as expressed on p9-8 because of the lack of sufficient data for this effort.

Presentation of COPECs NOAECs

p9-8, The objectives of a bioassay at MFA were to describe the exposure-response relationship for the range of concentrations and habitats to determine what concentration of the contaminant produced what response for the receptor. The bioassay effort was not just to determine the NOAEC because as we have now seen, the NOAECs were not readily obtained. The bioassay data should be further developed to show a relationship between the sample concentration for all contaminants and the response in the bioassay.

p9-8, **Storm-water Retention Pond**. The text states, "The lowest observed adverse effect concentration (LOAEC) (for sample SWRP-32) was at 12.5 percent sample concentration; therefore, the NOAEC is the 6.25 percent test concentration. The NOAEC concentrations are listed in table 9-4..." Were these concentrations actually measured from the sample? Do the concentrations represent sample SWRP-32? The concentrations at the other two sites shown,

SWRP-27 and 29 were above the NOAEL as determined in SWRP-32, however, there were many more samples that should be compared to these data. Additional data may indicate a concentration that is more representative of the actual effect.

Ditches and Channels.

p9-9, Sample SWNC-18 is listed as the only bioassay to have exhibited a "toxicity response" which means a positive hit, however the other sample results for organism and chem should be instructive for indicating a result of "no effects" and a suite of chemistry results. These data should be presented and evaluated as well. A lack of toxicity can be as important as direct toxicity if the tests were performed with logical experimental design. If the Navy believes that the observed manganese and copper concentrations may be a problem based on the comparison to "MFA aquatic life standards or criteria" a comparison should be made to the "literature history" or what is known about these two chemicals and impact to the appropriate bioassays. The comparison presented does not provide a "baseline" or "standard" for comparison.

In fact, there is a general misunderstanding by the Navy for the value of the "no response" data observed in the bioassays. In several parts of the document, e.g., Chapter 9, a "no effects" observation in a bioassay result (amphipod burial) is suggested to "lack sensitivity" when in fact the information should be further evaluated for the sediment relationships (for the amphipod tests) and chemistry levels in other test results (for acute bioassay results) to define those sediment and chemical "ranges" that do not produce a critical impact.

RISK ESTIMATION-SEDIMENT RECEPTORS

p9-11, Risk estimates should be based on bioassay information, not HQ estimates. The transfer of bioassay results to site specific risk does involve several assumptions that must be made including, the bioassay or test represents the mechanism of exposure as is found on the site for similar organisms, the test organism best represents the receptors found on the site, the results can be used to define the exposure-response for that particular organism and that this information can be transferred to the organisms and site of interest. This is the basis for performing bioassays during the Phase 2 risk estimation effort.

COPEC Risk Drivers.

p9-13, Although the cutoff point for HQs in this process is suggested to be 1.0, there are several incidence where values close to 1.0 should be included as well. The "algorithm" to describe the risk drivers is not an acceptable procedure because there is little basis for its use, it does not consider "hot spots" and the components are ambiguous. The "frequency of detection" depends, in part, upon the sample placement, the number of samples, and the laboratory effort to correct "matrix interferences." The frequency of detection is normally considered to be 100% otherwise a very good description of contaminant distribution based on sufficient sample numbers. The assumption that the frequency of detection term reflects the frequency of COPEC occurrence in an entire geographic area (p9-14 cited from Section 6.2) is intuitive, however the number of samples and placement of samples for MFA does not reflect a random placement and is therefore biased.

Further, the use of the HQ value is questionable because it is based on data that are highly uncertain. A better relationship would be the exposure-response "curve" or relationship that is

derived from bioassays. The use of the HQ is at best a screening number and should be used with caution in that process.

There are several questionable manipulations of these data that are unacceptable because they do not have a basis for the risk assessment. Table 9-6 shows the data that went into the "risk driver values" for the Eastern Dike Marsh where HQ values for 17 contaminants shown to be above 1.0 are "adjusted" using several manipulating "factors" to produce two "final" risk drivers. The range of HQs for this exercise was 1.15 up to 1211 with the "cutoff" point being between 63 and 44 to make it through the process as a risk driver. What is the basis for dividing the final risk driver value by the "total risk driver" value? This manipulation serves but one purpose and that is to reduce the number of contaminants of concern that produced an HQ identifying a potential problem. This manipulation of the HQs removes the importance of the biological components of the process because nothing is carried forward in these numbers to show the receptors, the pathways, and the overall impact to the receptors.

Spatial Distribution of COPEC risk drivers.

p9-15, The figures (Figs 9-1 through 9-10) are a beneficial means for displaying the potential risk "drivers" however, there are many contaminants that are not shown as the final list shown in Table 9-6 (for the Eastern Diked Marsh) incorrectly eliminates many of the contaminants that were shown to have HQs above 1.0 and are stated to present potential risk. Apparently the metal contaminants that were identified as COPECs were eliminated based on the suggestion that they are "ubiquitous" as reported in Section 5.2.2, which lacks justification for this decision. Tables 9-6 through 9-9 show the manipulation of numbers that eliminates contaminants from consideration that are observed in 30%, 70%, 20% with resulting HQs above 1.0; this is not acceptable.

Interpretation of Bioassay Data.

p9-16, The polychaete was suggested for testing to represent a food item for avian species rather than as a sediment dweller alone. The FETAX protocol was suggested as an evaluation of the possible effects from sediment and water on amphibians. Even with a dilution of 1 part sample water to 4 dilution water, the assay is useful as guidance for potential impacts from contaminated sediments.

Amphipod bioassays.

p9-18, The amphipod data should provide useful information on the potential impact to sediment dwelling organisms that provide food to wading birds and fish. I do not agree with the assumption on the Navy's part that the "reburial was an insensitive endpoint" as this aspect of the amphipod test indicates the "acceptability" of the sediment to the amphipod. Generally, if the amphipod shows high levels of reburial as was the case for these tests, the sediment is considered to be suitable for the amphipod, with the opposite (lack of burial) being meaning the sediment is not suitable for some reason e.g., grain size. Because of the high reburial (near 100% for all tests) the sediments in the various samples are considered to be adequate and should not be a major factor in any resulting mortality. Except for this observation in the amphipod testing, the questions about grain size was not addressed by the Navy (see p9-19) which is one shortcoming of the bioassay testing activity for this effort.

Polychaete bioassays.

p9-19, The polychaete tests were not suggested to be a test for mortality because in discussions with the Navy, these tests were known to be less sensitive than other organisms e.g., amphipods and were therefore, recommended for a measure of bioaccumulation for evaluation of food chain effects. A 30% reduction in growth is not a *de minimis* impact to polychaetes. As was suggested, the uptake of contaminants must be compared to the sediment levels of contaminants (in scatter plots) to demonstrate the levels of sediment contamination that potentially results in food chain impact to avian receptors.

Echinoderm bioassays.

pp9-19,20. The testing results produced sufficient data to define the levels of dilution that produced a significant impact and one that indicated no effect. Questions about the influence from stressors that are "non COPECs" are best addressed by very directed and specific bioassays such as Toxicity Identification Evaluation (TIE).

Fetax Bioassay.

p9-20, This bioassay, like the polychaete, was suggested for a specific reason that included a function for screening the sediment in the freshwater area of MFA, which was deemed a problem for the Navy in testing, because all of the proposed testing up to this point were estuarine or marine protocols and were thought to be inappropriate for this area. The FETAX was suggested to "determine if there was a general problem" with the freshwater sediments and to provide additional information about the potential impact to amphibians that was lacking. The mortality of 17 percent and a reduction in successful development (determined to not be statistically significant) have not been adequately incorporated into the risk characterization phase of this document.

Correlation of bioassay data with COPEC concentration.

p9-21, Table 9-11 is cited as providing a summary of correlations for effects, however the tabular information is lacking in detail. For instance, the correlation values should show the corresponding probability value and what is meant by the column heading, "Cluster Analysis and Comparison?"

Non-COPEC stressors.

p9-21, As previously stated, this aspect must be examined by very specific bioassays as we have seen with contaminants; concentrations alone do not definitively show impact to receptors.

Amphipod-based bulk sediment COPEC NOAECs.

9-22, These data are appropriate for determining the NOAEC, however, I don't remember that we agreed on the benchmarks shown, nor the sources. Were these benchmarks derived from amphipod studies?

Polychaete-based bulk sediment COPEC NOAECs.

p9-22, Again these data may be appropriate to show that no mortality was observed, however, where are the tissue data and the benchmarks for bioaccumulation?

Echinoderm-based NOAECs for Pore Water COPECs.

p9-23, Because of the test protocol, these test data are presented as a dilution series which provides the NOAEC for both development and mortality.

Risk estimate uncertainties.

p9-24, The Navy makes the statement, "While the bioassay results strongly indicate that sediments pose risk to benthic receptors, the lines of evidence do not converge very well on particular COPECs as the principle stressors and indicate that naturally occurring non-COPEC stressors may have a role in sediment toxicity." Up to this point, few of the bioassays have been performed in a manner to demonstrate the exposure impact relationship. The four listed "points" of uncertainty could be clarified by appropriate sampling.

RISK ESTIMATION - AVIAN RECEPTORS

Identification of Risk Drivers.

p9-28, This effort involved four steps, some of which are not logical and are therefore unacceptable. For instance, only HQs that accounted for at least 90% of the total HI are retained. The elimination of HQs and therefore, contaminants of potential concern using this process assumes that the data used for the various HQs is adequate and compatible when the data input varies quite a bit.

RISK ESTIMATION - MAMMALIAN RECEPTOR

p9-38, I would disagree with the count of HQs that indicate a potential problem as indicated by the Navy. Lead, selenium and total aroclor which the Navy did not include should be maintained along with arsenic, manganese, nickel, and zinc.

Chapter 10

RISK DESCRIPTION - SURFACE WATER

Risk description (for the various areas) is stated to have the objective, "...to present the risk assessment results in a way that is useful to risk managers in any site-wide remediation decision."

p10-1, **Risk Summary**. The aquatic setting at MFA includes a once natural system that has been altered for the most part including the elimination of tidal influence such that all isolated areas, the diked marsh and the storm-water retention ponds are greatly influenced by surface water flows. The Northern Channel has not been isolated, receives water from surface drainage systems, but is hydrologically connected to the Bay. The biological resources may have been altered dramatically as a result of these changes from the natural state such that there is little if any interchange or migration e.g., recruitment from the Bay to the stormwater retention ponds

and the fresh water marsh areas. Because of these alterations, the overall complex of organisms and plants may be quite different than that of the Bay and surrounding tidal marshes.

p10-2, **Receptors at Risk**. This information may have suffered from the limited effort and the lack of seasonal coverage. Apparently a previous contractor for the Navy performed a limited survey of the areas and produced a very cursory report of the biological resources.

The focus of the evaluation for receptors was, "...direct contact with surface water ...the bioassays and HQs based on water concentrations." Surface water bioassays included algae, invertebrates, and fish which were proposed for the purpose of determining the direct relationship between exposure and response of the organism. The invertebrates and the fish bioassays showed a significant impact after being exposed to the appropriate media and the algae tests did not result in a significant impact. This information should be summarized by displaying concentration and response scatter plots.

The avian and mammalian receptors were evaluated by gathering information to develop toxicity reference values (TRVs) for all pathways and chemicals. Apparently, from the HQs estimated for surface water, there was little if any risk predicted from this pathway. The estimation of "proportions" of total dose for the purpose of elimination from further consideration is not appropriate for this process. Surface water must be considered as part of the ingestion pathway for avian species, otherwise the evaluation is incomplete.

p10-3, **COPECs driving risk**. These data for Table 10-2 originated from Tables 8-1 through 8-3 showing a close representation, however there are three values that should be included as follows: lead with an HQ of 0.95 and 0.946 and nickel with an HQ of 0.975 all of which are too close to 1.0 to be eliminated. Also, chromium in the Eastern diked marsh had a HQ of 1.12 and silver at 1.64 which are certainly above 1.0 and should be included.

p10-4, **Summary of Uncertainty**. Table 2 is incomplete by not listing several major possible areas of uncertainty for this assessment such as lack of sufficient sample numbers, sample coverage, detection limits, sufficient and appropriate chemical analytes, sufficient bioassays, experimental design, data presentation and data analyses.

p10-4, **Problem Formulation**. Yes, the "qualitative" biological survey is a short coming of the overall process that could have been avoided. There is little reason, again, for any uncontrollable uncertainty between the measurement and assessment endpoint since both of these are choices of the risk assessor.

p10-5, **Analysis**. What is the relationship of the sediment contamination to the water column concentrations? Many of these uncertainties (seasonal variations, random sampling in addition to a focused effort, dissolved concentrations for water column invertebrate and total for water column vertebrates) could have been avoided by a better experimental design.

p10-5, **Risk Characterization**. I believe that there are only two contaminants that are antagonistic to each other, selenium and mercury. All others are most likely additive based on the literature. The effort at MFA, like all other sites would be successful if a sufficient effort were completed for single species efforts rather than think that populations or communities and much less ecosystems are being assessed. The material presented for the MFA assessment does not address any concerns other than single species.

Weight of Evidence.

p10-6, **Sufficiency and quality of data**. The quality of these data may be sufficient, but there is doubt that the coverage and the right kind of data was adequate. More bioassays should have been performed, better chemical analyses to avoid "matrix interferences" should have been emphasized, PCB congeners should have been emphasized in more locations, a better effort for defining the exposure-response relationships should have been emphasized for sediments, biouptake, growth and reproduction tests.

p10-7, **Corroborative information**. The most appropriate way of corroborating the results of bioassays and other toxicity evaluations is to perform confirmatory tests that demonstrate that the results obtained are in fact accurate and representative. Broad, complicated field exercises are generally ill planned and thought out and do little except waste time and money.

There is too much emphasis being placed on the "limited biological community" at MFA. The environment at MFA has been altered physically and chemically, this risk assessment effort focused on the chemical impacts and, if successful, should be quite useful to enhance the biological resources at MFA with reductions in the chemical impact to these resources. There was insufficient data collected to suggest that, "The observed reduction in growth alone would not be expected to cause the absence of the number of organisms observed during the biological survey."

p10-8, **Evidence of causality**. **Strength** - lack of sufficient samples in Northern Channel and the Eastern Diked Marsh. **Consistency** - what is the basis for making the statement, "...there was an inconsistent response to similar levels of COPECs in the storm-water retention pond bioassays?" Where are the data to show that the tests were performed at the same level of contamination? **Specificity** - perhaps there were insufficient tests performed to verify the specificity? **Temporality** - lack of sufficient data were obtained to evaluate this criterion. **Presence of a biological gradient** - the test gradient was demonstrated in the tests with serial dilutions; there were insufficient tests performed to evaluate a spatial gradient, however, when the contaminants are examined for presence and absence at HQs above 1.09, there is a definite trend from the source (Lindberg Ditch) through the Eastern Diked Marsh into the Storm water retention ponds. **Plausibility** - this criterion is met by metals and growth endpoint. **Coherence** - the HQs should not be expected to correlate with the bioassays, at least by the methods presented. There may be other influences that affected the bioassays, however the quality and type of data used for the HQs do not suggest that a correlation would occur. **Experimental evidence** - there were several problems with experimental design. **Analogy** - again, not sufficient data to evaluate this aspect.

INTERPRETATION OF ECOLOGICAL (sic) SIGNIFICANCE

This is a misconception about ability to achieve a level of evaluation that was not planned for, data were not collected, and the information is not available based on the material presented.

p10-10, **Nature of effect**. The "nature of effect" is difficult to evaluate because the information gathered and presented do not relate to a true "ecological" risk assessment. The reduction in growth in a crustacean and a fish are significant in that no data were collected for zooplankton, including crustacean food items of fish that might inhabit the Storm water retention ponds if they were not isolated from the Bay. Although there was no "statistically" significant impact to the

alga tested, other indigenous species may be impacted. There were no evaluations of epibenthic organism such as bacteria and microcrustacea and other orders of invertebrates (Nematoda) that serve as food items for larval fish. There may not have been an adequate gathering of data to fully evaluate the potential impacts to "the energy cycle" at MFA.

p10-11, **Magnitude of effect**. Again, where are the data presented to show that there was a range of responses at similar concentrations? The reference to a significance being attributed to a 25% reduction (Suter says 20%) is not appropriate to use for a general or generic benchmark. Barnthouse et al (1986) suggested this level of impact for game fish, probably a centrachid, not a top minnow as Menidia, a different fish and different guild, so this comparison may be inappropriate or irrelevant. I would agree that the HQs should be used very cautiously only for screening and are questionable for, "...applying to the determinations of potential impacts to the community."

Again, there is too much emphasis on the "limited number of surface water organisms" noted during the biological survey. The "survey" was taken one time during the annual cycle, it was a qualitative effort and given sufficient data collections and testing, contaminants may be the cause for this apparent paucity of biological life.

p10-12, **Spatial extent of effects**. What are the characteristics of the active fate and transport system at MFA? There are sufficient data to show that the contamination is equally as bad in the Eastern Dike Marsh as the Storm Water Retention Ponds which are less than the Lindberg Ditch was before it was remediated (look at that data). The surface waters exhibit a pattern of contaminants from the Eastern Dike Marsh to the Storm Water Retention Ponds that suggests the path of distribution.

p10-12, **Duration of effects**. What are the bases for stating that the concentrations of COPECs will not remain constant over time? From the above paragraph, the last sentence is cited, "It is possible that other factors such as sediments, may be playing a role in the level of COPECs in the surface water as well..." And if "natural processes" such as adsorption, precipitation, and degradation would be expected to reduce the level of COPECs in the surface water over time, where would these contaminants sink, the sediments in a closed system such as the Storm Water Retention Ponds?

p10-12, **Likelihood of Recovery**. These statements are not based on any logic presented in the report, however, we remain optimistic.

The bioassays that were performed in a serial dilution, the echinoderm development and mortality showed a definite relationship of greater impact with increasing concentration.

The amphipods showed a response that if presented in a scatter plot would most likely show a similar trend, the FETAX showed a response to contaminant strength. The problem in this area of "proof" is the lack of adequate data due to poor experimental design and to the lack of proper data presentations, not the lack of causality.

Chapter 11

RISK DESCRIPTION -SEDIMENTS

RISK SUMMARY

p11-1, **Receptors at risk**. The bioassays and direct measurements of sediment chemistry are independent measurements to determine the potential risk to receptors. The HQs as presented have already been discussed and shown to have little relevance to the bioassay results, especially as a predictive tool for bioassay results. Again, this effort as presented does not offer any data for the evaluation of populations in a risk assessment effort.

Summary of receptors at risk based on hazard quotient.

p11-2, Receptors inhabiting sediments of the Eastern Diked Mash and the ditches and channels have a relatively high potential to be adversely affected by sediment COPECs. Receptors inhabiting sediments of the storm-water retention ponds have a relatively moderate potential to be adversely affected by sediment COPECs. Receptors inhabiting sediments in the Western Diked Marsh have a relatively low potential to be adversely affected by sediment COPECS.

Receptors at risk based on pore water chemistry and sediment toxicity tests.

Based on the bioavailability information and exposure route information, infaunal animals for which pore water uptake is a major exposure route have the most potential to be adversely affected by MFA sediments. Relatively low COPEC bioaccumulation occurred and with the exception of one test where growth was reduced by 30 percent, the worm generally showed no toxic effects when exposed to sediments, This information suggests that 1) COPECs associated with particulate material may be bioavailable at low levels, i.e., not extremely toxic, and 2) that ingestion may be a minor exposure route compared to pore water uptake. The FETAX bioassay on sediment from the Eastern Diked Marsh showed a significant reduction in embryo survival. This limited information suggests that amphibians (and possibly other organisms) that inhabit the sediment-water interface may be at risk due to toxicants leaching from sediment.

COPECs driving risk to benthic receptors.

p11-3, I do not believe that the indices, based on bulk sediment COPEC concentrations is the appropriate way to integrate the risk data for this site. This is based on the opinion that sufficient data were not collected because of poor sample coverage and sample numbers with limited analyte data. The bioassay data should be developed to present exposure response relationships.

Relevance of measurement endpoint information

p11-4, The HQ for bulk sediment and pore water are not measurement endpoints as discussed in Chapter 8 comments.

p11-5, The bioassays are not considered measurement endpoints. The bioassays were not performed in this study to "discern between adverse effects caused by COPECs and adverse effects caused by naturally occurring factors, such as high levels of ammonia and hydrogen sulfide" because "laboratory" controls were not designed into the testing.

Confidence in risk estimate.

p11-6, Geographic areas - There were not sufficient samples collected from all areas of concern. This is true for the amphipod samples, the FETAX samples, the echinoderm samples all except the echinoderm test were not performed in a manner to obtain a response to a gradient of chemical exposures representing a gradient of risk estimation.

Confidence in chemical stressors posing the risk.

p11-7, The pore water, the amphipod, the FETAX and others perhaps if presented appropriately would show a response to the gradient.

Confidence in benthic receptors at risk.

p11-7, The distribution of COPEC concentrations in the bulk sediment chemistry data set **may** represent the spatial distribution of COPEC concentrations at MFA. The faunal composition in MFA sediments may not be expected to be similar to tidally influenced mud flats that are hydrologically connected with San Francisco Bay surface waters. The benthic fauna at MFA was not adequately characterized and the total numbers and composition is not well described.

ECOLOGICAL SIGNIFICANCE OF RISK ESTIMATES

p11-8, The Navy **cannot** assert that the MFA sediments support a benthic community that is generally similar to the environments of estuarine tidal flats of San Francisco Bay. The Navy has maintained elsewhere in this document that the "benthic fauna is of low diversity and contains relatively low numbers of organisms."

p11-9, The Navy did not measure, collect data, nor evaluate the microfauna (0.1 mm) and therefore cannot suggest that the organisms less than 0.5 mm may be dominant in terms of total metabolism.

Nature of adverse effect.

p11-10, The Long et al, 1995 data set can guide the Navy in predicting impacts and resultant risk, but risk cannot be determined from these data.

Magnitude of adverse effects.

Bioassays are the only means by which the magnitude of impacts from chemical contaminants can be determined. HQ estimates cannot be used because of the high level of uncertainty. Bioassay data collected for MFA were inadequate to provide detailed extrapolation to estimate the magnitude of effects because of poor experimental design, not because of the suggested inability of bioassays. Actually, the results of bioassays (polychaete growth, amphipod survival, echinoderm survival and development and FETAX development and survival) indicate that MFA sediments may pose a moderate to high potential to adversely affect some fauna through exposure to bulk sediments and/or sediment pore water.

Effect on endpoint receptor populations.

p11-11, Polychaete tests were intended to provide an assessment of food chain impact, not an assessment to invertebrate survival. Because of the design, polychaete populations could not be

evaluated i.e., no reproductive measurements. *Eohaustorius estuarius* showed moderate to high levels of mortality when exposed to MFA sediments. What data supports the statement, "...numerically dominant amphipod, *Ampelisca abdita*? The sediment pore water bioassays indicate that sensitive biota, for which pore water is a major exposure route, will be severely affected by MFA sediments. The FETAX bioassay suggests that organisms with early life stages that inhabit the sediment-surface water interface could experience adverse effects from toxicants leaching from the sediments into the sediment-surface water interface. This is especially true if the dilution of 1 part sample to 4 parts of dilution water are considered when interpreting these data. Essentially, the concentration could be four times greater than reported because of the dilution factor.

Community effects.

p11-12, The first paragraph is not based on any information that was gathered from MFA, in fact, there is much speculation without foundation, "...theoretically, these species may be displaced over time by resident species, as the habitat evolves from disturbed state to an equilibrium, density-dependent state." There was insufficient data collected to evaluate community effects.

Spatial extent of adverse effect.

The spatial extent of significant impacts is one of the primary goals of risk characterization. The use of Thiessen polygons helps to display the risk estimates in a very productive manner.

Duration of Adverse effects.

p11-13, Yes, I agree, the duration of adverse effects will continue until exposure of benthic receptors is reduced. This is true whether or not the benthic receptors at MFA are similar or different than SF Bay mud flats.

Likelihood of recovery.

Sedimentation rates are unknown at this time; why would they change from past rates? Much of what is written here is without site specific basis.

RISK SUMMARY FOR AVIAN AND MAMMALIAN RECEPTORS

Receptors at risk.

p11-15, Both avian and mammalian receptors may be at risk as a result of exposure to sediments at MFA. The great blue heron appears to be at greater risk than other receptors even when "...only indirect exposure to sediment contamination (via the food chain) was assumed" without considering the incidental ingestion of sediment.

There is nothing provided to suggest that incidental sediment ingestion contributes little to the total dose for the black-necked stilt, mallard duck, and the salt marsh harvest mouse. Even when stating, "Additionally, the incidental sediment ingestion rates vary between receptors" the Navy states, "In considering the percent contribution of incidental sediment ingestion to the total risk, the ingestion rates for two receptors were assumed to be the same."

Sampling design.

p11-18, The purpose of the sampling design was to determine the maximum concentration rather than the mean.

Adherence to sampling and analysis protocols.

p11-19, Some sample fractions for PAHs, OC pesticides/PCBs and TPH-extractable were not met for one of two planned samples and 7.1 percent of the PAH compounds from surface water, sediment elutriate, and sediment pore water samples.

Causal relationship between COPECs and effects.

p11-21, Whether or not any field testing was performed, the laboratory test results are clearly in support of causal relationships. The bioassay should be considered highly supporting evidence that there is some relationship between the contaminants and the responses of the test organisms.

INTERPRETATION OF ECOLOGICAL SIGNIFICANCE FOR AVIAN AND MAMMALIAN RECEPTORS

p11-23, The interpretation of ecological significance should rely on the data generated from the risk assessment rather than professional judgment. Consideration should be given to the nature and magnitude of effects and spatial and temporal patterns of effects. The estimate of recovery potential must be a very focused effort involving the evaluation of the remedial options.

The nature and magnitude of effects should be a result of the relationships developed from the bioassays performed on the site i.e., exposure-response relationships, how extensive the significant effects concentrations extend over the site. Also, the extent of the severity of impact, whether or not the impacts were at the individual, the population or community level all of which indicate the magnitude of effects and provide some potential for estimating the recovery time.

The bioassays that were performed in a serial dilution, the echinoderm development and mortality showed a definite relationship of greater impact with increasing concentration.

The amphipods showed a response that if presented in a scatter plot would most likely show a similar trend; the FETAX showed a response to contaminant strength. The problem in this area of "proof" is the lack of adequate data due to poor experimental design and to the lack of proper data presentations, not the lack of causality.