UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
REGION IX
75 Hawthorne Street
San Francisco, California 94105

August 14, 2018

Lawrence Lansdale, Environmental Director
Base Realignment and Closure Program Management Office
U.S. Department of Navy
33000 Nixie Way
San Diego CA 92147

Dear Mr. Lansdale:

The U.S. Environmental Protection Agency (EPA) has reviewed the Navy's draft *Parcel G Removal Site Evaluation Work Plan, Former Hunters Point Naval Shipyard*, San Francisco, California, June 15, 2018 (*Work Plan*). EPA acknowledges that the need for this Work Plan is due to a highly unusual and serious situation. It is for this reason that EPA and California Department of Toxic Substances Control (DTSC) wrote the Navy in 2016 to reiterate our agreement with you that no further transfers of property should occur until we fully investigate our concerns regarding all data collected by Tetra Tech EC Inc. At the same time, EPA stressed that retesting in locations of greatest concern should take place as soon as possible.

Over the last two years, my staff has worked daily with the Navy, DTSC, and California Department of Public Health (CDPH) to review data and plan for retesting. Where regulatory agencies disagree with the Navy, we are guided by our Federal Facility Agreement (FFA) and the vast body of site cleanup decisions and past practices that protect public health and the environment for this community. To ensure the integrity and credibility of decisions in this unprecedented situation, we also apply established national Superfund laws, regulations, and policies as they relate to a site like this. At Parcel G, EPA and our State partners have worked diligently to lay out a scientifically driven retesting strategy that, if followed, is designed to provide confidence to the regulators and all the stakeholders when the site would be suitable for redevelopment.

Attached are EPA's review comments on the June 2018 *Work Plan* for retesting the potentially radiologically impacted portions of Parcel G, which includes storm and sewer lines, building site soil areas, and buildings. These new comments largely repeat the comments EPA already submitted in March 2018[1] in response to the Navy's previous February 2018 draft Work Plan for retesting parcels in general. They are also consistent with recommendations that EPA has given the Navy since Fall 2016 to retest expeditiously in areas of greatest concern.

---

1 EPA's March 2018 Comments are at this link: https://semspub.epa.gov/work/09/100009179.pdf

EPA's main concern is that the approach described in the *Work Plan* does not reflect the Parcel G proposed retesting approach that all agencies (EPA, DTSC, CDPH) submitted in our March 2018 comments. EPA's review of previous data collected by Tetra Tech EC Inc. in six parcels found potential signs that 90 to 97% were unreliable, as we wrote in December 2017 and March 2018. The Navy has agreed to retest all of the survey units where Tetra Tech EC Inc. did previous radiological work. Our review of the *Work Plan* identified the following key elements from our March 2018 comments that are missing from the *Work Plan*:

1. **Initial Testing Stage**

A new Superfund investigation typically begins by collecting historical information and reviewing existing data. In Parcel G, the Navy, EPA and other agencies reviewed site history and signs of potential falsification in the existing data. The next task is typically testing in the field first in the areas of greatest concern. In Parcel G, using a statistical approach designed to protect public health, EPA and our State agency partners selected the 33% of trench units most likely to show concerns based on our reviews. This approach ensures a high confidence that the trench and building site survey units meet Remedial Goals (RGs).[2]

EPA agrees with the Navy's plan in Phase 1 to excavate and test for contamination[3] in these 33% of trenches with the highest concern. EPA also agrees that even if no contamination is found in Phase 1, core sampling (without excavation) to test material under the surface must still be performed in the remaining 67% of trenches (the Navy's "Phase 2"). However, the draft final *Work Plan* must also include these elements:

- Core sampling must be done in no fewer than the number of systematic locations determined from a statistical evaluation consistent with the practices described in the *Multi-Agency Radiation Survey and Site Investigation Manual* (MARSSIM), a widely-accepted approach. Each core sample is considered to be a single systematic sample location, even though multiple depths within the core may be analyzed.

- In addition, even if no contamination is found in Phase 1, in the remaining 67% of trenches, no less than 100% of the surfaces must be scanned, after removing existing asphalt and cover material.

---

[2] The *Work Plan* references EPA's 2014 guidance on radiation risk assessment at Superfund sites. Please understand that the governing cleanup standards at Parcel G are the RGs established in the 2009 Parcel G Record of Decision (ROD). Only areas that are demonstrated to comply with these requirements will be eligible for EPA approval and release.

[3] Contamination is defined as radionuclide concentrations above the RGs in the 2009 Parcel G ROD, after subtracting reference background, if not proven to be Naturally Occurring Radiological Material (NORM) or anthropogenic background, e.g. Cs-137 fallout.

2. Followup Actions Stage

As a next stage, the regulators and the responsible party would typically review the initial stage testing results to determine next steps. Our March 2018 comments already laid out next steps, depending on the extent of contamination. The following elements of these steps are missing and must be included in the draft final *Work Plan*:

- If any single sample in Phase 1 shows contamination, then core sampling and scanning (without excavation) would no longer be sufficient by itself to demonstrate protection of human health and the environment. In this scenario, the entire remaining 67% of the trenches, i.e. a total of 100% of trenches, must also be fully excavated and tested.

The draft final *Work Plan* must follow a similar staged approach for soil at former building sites. Our attached comments also address testing related to buildings, other aspects of soil, and reference background.

We may supplement our comments after evaluating any public comments received by the Navy on the draft *Work Plan*. EPA then expects a draft final *Work Plan* that reflects our final comments, so EPA and the State can swiftly approve it as final. Without the requested changes, the approach will not provide the necessary confidence level to establish when Parcel G would be suitable for redevelopment, and EPA may invoke the dispute resolution process described in the FFA.

EPA will continue working with the Navy, DTSC, and CDPH to ensure protectiveness, transparency, accountability, and substantive public involvement. Please contact me at 415-972-3144 or Herrera.angeles@epa.gov if you would like to discuss these issues further, or you may contact John Chesnutt, who leads the technical team, at 415-972-3005 or chesnutt.john@epa.gov.

Sincerely,

Angeles Herrera
Assistant Director, Superfund Division
Federal Facility and Site Cleanup Branch

Attachment: EPA's comments to Navy on its June 2018 draft *Work Plan*

cc: Janet Naito, California Department of Toxic Substances Control
Anthony Chu, California Department of Public Health
Barbara Garcia, San Francisco Department of Public Health
Nadia Sesay, Office of Community Investment and Infrastructure
Terry Seward, Regional Water Quality Control Board

EPA Review of the June 15, 2018, Draft *Parcel G Removal Site Evaluation Work Plan*, Hunters Point Naval Shipyard, San Francisco, California
EPA Review dated August 14, 2018

GENERAL COMMENTS

1. **Executive Summary; Section 2, Conceptual Site Model; and other sections:** The June 2018 draft *Parcel G Removal Site Evaluation Work Plan* ("Work Plan") acknowledges many aspects of the 2008 Conceptual Site Model (CSM)[1] for storm drain/sewer lines that is cited in the Radiological Removal Action Completion Report (RACR) the Navy produced for Parcel G and other parcels. This 2008 CSM states that contamination could have come from any leaks in storm drain/sewer lines, which could have been a result of many factors that could have occurred at any locations along the lines. (See General Comment # 21 in the U.S. Environmental Protection Agency [EPA] December 2017 comments on the radiological data evaluation for Parcels B and G).[2]

   The EPA, State of California Department of Toxic Substances Control (DTSC), and the California Department of Public Health (CDPH) found that the original test results from Tetra Tech EC Inc. are unreliable. Therefore, we are relying on the original 2008 CSM that states that "The potential for materials to migrate from piping/ and manholes into the surrounding soils is significant." The Executive Summary and Section 2, Table 2-1, "Uncertainties" section lists factors that could result in "Lower potential for radiological contamination than originally described in historical CSMs." While some of these factors could theoretically affect the extent of contamination potentially left behind by Tetra Tech EC Inc., until new reliable testing results are available, the 2008 CSM stands. This CSM was the basis for the EPA March 2018 comments on the Navy's February 2018 draft Work Plan for retesting any parcels.

   In addition, the Executive Summary and Table 2-1 also refer to anthropogenic fallout as a potential source for Cesium 137 (Cs-137). Previous radiological work at the Hunters Point Naval Shipyard (HPNS) did not apply a reference background value for Cs-137 except in Parcel E-2.[3] While the EPA has no objection to collecting new reference background data for Cs-137, please refer to this comment EPA previously submitted December 29, 2017, to the Navy about Cs-137 contamination due to Navy activity at Parcel G: "the Navy has found radiological contamination in portions of Parcel G, such as in the southeastern corner (associated with the buildings and the "peanut spill") and in the sewers along Cochrane Street due to previous testing during the Phase I through Phase V Radiological investigations/cleanups. The 2004 HRA [Historical Radiological Assessment] indicates that Cs-137 was found at high concentrations in sediment from a manhole along Cochrane

[1] *Navy Memorandum for the Record: Conceptual Site Model for the Removal of the Sanitary and Storm Sewers at Hunters Point Shipyard*, December 17, 2008
[2] EPA Review of Draft *Radiological Data Evaluation Findings Report for Parcels B and G Soil, Former Hunters Point Naval Shipyard, San Francisco, California*, Comments Dated December 2017. Link at EPA website: https://semspub.epa.gov/work/09/100006302.pdf
[3] ERRG, 2011.

Street."[4] The HRA documents that the Navy used Cs-137, resulting in liquid waste that resulted in releases in Building 364 in piping, sinks, and the "peanut spill" behind the building. The HRA also documents in Table 5-1 that the Navy had 5 radioactive licenses with the Atomic Energy Commission for Cs-137, one for a quantity of 3,000 Curies and a separate quantity of 20 Curies of Cs-137. Two licenses indicate that Cs-137 was in sources. In some cases, the Navy made their own sources with Cs-137.

Please add to the Executive Summary text that Parcel G has contained Cs-137 contamination due to the Navy's activities. In Table 2-1, "Potential Source Areas" Section, please revise the text to indicate the sources related to Cs-137.

As a result of the above history, until receiving any evidence to the contrary, the underlying assumption should be that new comprehensive testing is necessary and that Cs-137 found in new testing could be due to Navy contamination. The regulators are open to evidence for an alternative CSM, such as new reliable data about the extent of contamination found after excavating the trench units (TU's) most likely to have contamination. Contamination is defined as radionuclide concentrations above the RGs in the 2009 Parcel G Record of Decision, excluding Naturally Occurring Radiological Material (NORM) or anthropogenic background. Excavation and testing of the soil survey units with the greatest likelihood of contamination is an important step toward testing the validity of the original CSM. Please ensure future versions of the Work Plan and the updated Master Sampling and Analysis Plan (SAP) address EPA's assumptions about the CSM.

2. **Executive Summary; Section 3, Soil Investigation Design and Implementation; and other sections:** The June 2018 Work Plan does not include necessary elements of the retesting proposal presented in EPA's prior comments in March 2018. Based on the original 2008 CSM, EPA, DTSC, and the CDPH proposed in March 2018 a scientifically driven retesting strategy that, if followed, is designed to provide confidence to the regulators and the public when the site would be suitable for redevelopment. The details appear in EPA's attached March 2018 comments.[5] In addition, attached is a statistical review of the June 2018 Work Plan. For example, the Work Plan does not provide information about the path forward in a scenario in which contamination is found anywhere within the Phase I TUs or Survey Units (SUs). EPA stated in its March 2018 comments that if contamination is identified in any of the initial 33 percent (%) of TUs, then all the TUs in Parcel G (100%) will require excavation and testing. Similarly, for building site SU's, if contamination is identified in any of the initial 50% of SUs then all the similar units in Parcel G (100%) will require excavation and testing. Please revise the Work Plan to include this requirement. Similarly, Figure 3-2, Performance Criteria for Demonstrating Compliance with the Parcel G ROD – Soil, does not include a step in the logic diagram for the next steps to be taken if Ra-226 exceeds the RG (1.0 picoCuries/gram above background). Please revise Figure 3-2 to

[4] EPA Review of Draft *Radiological Data Evaluation Findings Report for Parcels B and G Soil, Former Hunters Point Naval Shipyard, San Francisco, California,* Comments Dated December 2017. Link at EPA website: https://semspub.epa.gov/work/09/100006302.pdf

[5] EPA's March 2018 comments are available at this link on the EPA website: https://semspub.epa.gov/work/09/100009179.pdf

include a complete logic diagram demonstrating actions that will occur if Ra-226 is found to exceed the RG in any sample.

3. **Executive Summary; Section 3, Soil Investigation Design and Implementation; and other sections:** The Work Plan proposes including cleanup criteria that are not documented in the Parcel G Record of Decision (ROD). The following sections contain language regarding additional cleanup criteria at Parcel G that are not documented in the Parcel G ROD and therefore do not meet the statutory requirements of the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) and the National Contingency Plan (NCP) of 40 CFR §300.430 Remedial investigation/feasibility study and selection of remedy:

    a. The Project Purpose section of the Executive Summary, states, "Portions of soil or structures at Parcel G that are not compliant with the RAO [Remedial Action Objective] specified in the Parcel G Record of Decision (ROD)] will be evaluated for protectiveness based on the United States Environmental Protection Agency's (USEPA's) current guidance on Radiation Risk Assessment at CERCLA Sites, Radiation Risk Assessment at CERCLA Sites (USEPA, 2014) [Radiation Risk Assessment at CERCLA Sites]." At this stage of the CERCLA process, the cleanup goals have already been legally established. A new Radiation Risk Assessment is ordinarily only performed as part of a Five-Year Review to evaluate whether or not the original RG's are still protective. EPA has separately recommended that the Navy conduct this review, and, if any of the RGs are found to be no longer protective using the most current risk calculators, propose amendments to the Parcel G ROD to ensure protectiveness. For the current work plan, however, the current RGs still govern the cleanup and if any material is found on Parcel G that exceeds the RGs established in the Parcel G ROD for the ROCs, excluding naturally occurring and anthropogenic background, the material should be removed and disposed of in accordance with the ROD and other applicable laws and regulations.

    b. The Executive Summary, Phase I discussion states, "To the extent practicable, soil with ROCs [radionuclides of concern] at concentrations above the RGs [remedial goals] will be evaluated further using USEPA's current guidance on Radiation Risk Assessment at CERCLA Sites."_As stated above, pursuant to the ROD, the remedy at Parcel G requires that "[b]uildings, former building sites, and excavated areas will be surveyed after cleanup is completed to ensure that no residual radioactivity is present at levels above the remediation goals" and "[e]xcavated soil, building materials, and drain material from radiologically impacted sites will be screened and radioactive sources and contaminated soil will be removed and disposed of at an off-site low-level radioactive waste facility."[6]

    c. The Data Evaluation and Reporting states, "If the investigation results demonstrate that site conditions are not compliant with the Parcel G RAO, then the data will be evaluated to determine whether site conditions are protective of human health using

[6] Parcel G ROD, pg. 43.

USEPA's current guidance on Radiation Risk Assessment at CERCLA Sites (USEPA, 2014). A removal site evaluation report will be developed to include recommendations for further action." EPA Directive 9200.4-40 was issued as guidance only and, as such, is not a regulatory requirement or a ROD-established cleanup level for the Hunters Point Naval Shipyard (HPNS) site in accordance with the CERCLA process as promulgated in 40 CFR §300.430. At Parcel G, the ROD has already established cleanup goals that govern the remedy. Please revise these sections of the Work Plan to state that only areas that are demonstrated to comply with the Parcel G ROD requirements will be eligible for Regulatory Agency approval and release.

4. **Executive Summary and Section 3.4.4, Phase I Trench Unit Investigation:** This section states that TUs will be over-excavated (i.e., excavated outside the estimated previous boundaries of the sidewalls and bottom), and will be gamma scan surveyed and sampled ex-situ (i.e., on a Radiation Screening Yard). The Work Plan Table 3-1, Phase 1 soil Trench Units indicates that the sidewalls and floor will be combined into one survey unit. The Navy's proposal to excavate all soil beyond the previous boundaries will be more protective than EPA's March 2018 proposal because more material will be excavated and tested instead of only systematic samples. In addition, scanning this material ex-situ will give more reliable results that scanning in-situ (i.e., in the trench itself). Therefore, EPA agrees with the Navy's alternative proposal to address the potential for contamination to remain in the sidewalls and bottom of the trenches. However, please revise the Work Plan to specify that in the event that an exceedance above any of the ROD ROC RGs is identified in the ex-situ scanning, the Work Plan should require in-situ investigation, i.e., the sidewalls and floor of the associated trench be scanned and systematic samples should be collected and analyzed inside the trench to identify where contamination may still be present. Furthermore, please revise the Work Plan to specify that the source trench will not be backfilled before confirming if an exceedance is found in excavated material. If an exceedance is found, then the trench will not be backfilled until the in-situ scanning and sampling is done to identify the location of the exceedance and excavation of contamination is completed.

5. **Executive Summary and Section 3.4.5, Phase 2 Trench Unit Design:** Because the surface of the trench is the location closest to potential residents, EPA recommends treating the surface over each former trench or survey unit as a new soil surface survey unit to be tested using an approach similar to that used in previous HPNS radiological investigation Work Plans and in MARSSIM. This means that after removing the asphalt and any other cover material, 100% scanning and systematic sampling should be conducted. The number of cores must be no fewer than the number of systematic locations determined from a statistical evaluation consistent with the practices described in MARSSIM. Each core location is considered to be a single systematic sample location, even though multiple depths within the core may be analyzed. In the past, 18 samples has been used as a default, but this number should be calculated based on the variability in the data actually collected, which may result in a total number higher or lower than 18. These calculations should use the variability in the sample results obtained from the new background study. Please revise the Work Plan to specify the number of locations for core sampling locations must be determined as described previously in EPA's General Comment # 20 in its March 2018 comments.

6. **Executive Summary and Section 3.4.5, Phase 2 Trench Unit Design, Page 3-7:** The text does not describe the percentage of land area for Phase 2 trenches that will receive gamma scanning. The Parcel G trenches should be treated as MARSSIM Class 1 trenches, as in previous HPNS radiological Work Plans, because of the CSM. The EPA stated in its March 2018 comments, "To address the potential exposure to future residents, 100% surface scans would be required. The Navy must first remove any asphalt cover and any imported fill that may have been used to achieve the desired grade, i.e. not part of backfill that potentially came from an area excavated by Tetra Tech EC Inc. Any locations where scan results exceed the investigation level would require collection of biased samples." Please revise the Work Plan to reflect this step.

7. **Section 1, Introduction:** This section states that a separate *Sampling and Analysis Plan* (SAP) will be prepared for the investigation at Parcel G, however the SAP has not yet been provided for review. The revised and updated SAP should be issued for review by the Regulatory Agencies prior to initiation of work at Parcel G. Information provided in the Work Plan and the SAP and any other supplemental documents (e.g. any Task Specific Plans) should incorporate all of the technical, as well as quality control (QC) requirements for sample collection and analysis, data validation, assessment and reporting, along with copies of standard operating procedures for all of these processes. The technical information should include the method number, calibration information and quantitation parameters. The QC information should include daily/weekly efficiency, energy and background checks as applicable; and results for matrix spikes, duplicates, blanks, Laboratory Control Samples (LCS) samples, tracers (alpha spectroscopy), and the following method-specific parameters:

   Gross alpha/beta Scans for Buildings Scan minimum detectible concentrations (MDCs) are below Investigation Levels for all radionuclides of concern (ROCs)

   Gamma Scans, Gross alpha/beta Scans Scan MDCs are below the Investigation Levels for all ROCs

   Gamma Spectrometry Static measurements or laboratory analysis
   - Sample results should include all radionuclides detected along with count times, result, counting error, and isotope specific MDCs
   - Demonstration that radionuclide-specific MDCs that are 10% of the ROC remedial goals (RGs) can be achieved.
   - A copy of the gamma spectrometry analysis library

   Alpha Spectrometry (See more detail in comment below)
   - All Uranium and thorium isotopes by alpha spectroscopy for samples with elevated Ra-226, count times, results, counting and total propagated uncertainty, MDC, tracer recovery
   - Demonstration that the Uranium (U)-234, U-235, U-238, Thorium (Th)-230, and Th-234 MDCs at 10% of the Radium (Ra)-226 RG can be achieved.

In summary, please ensure the Work Plan and SAP include all the specifics describing all radiation surveys, sample collection and analysis technical and QC requirements as described above. In addition, due to significant public interest, we recommend that the draft SAP be made available to the public for comment.

8. **Section 3.1, Data Quality Objectives, Step 5 – Develop Decision Rules, Page 3-1 and Step 7 – Develop the Plan for Obtaining Data, Page 3-2; and Section 4.1, Data Quality Objectives, Step 5 – Develop Decision Rules, Page 4-1:** The decision rules are not consistent with the EPA March 2018 comments and the requirements of the Parcel G ROD, which states, "Buildings, former building sites, and excavated areas will be surveyed after cleanup is completed to ensure that no residual radioactivity is present at levels above the remediation goals. Excavated soil, building materials, and drain material from radiologically impacted sites will be screened and radioactive sources and contaminated soil will be removed and disposed of at an off-site low-level radioactive waste facility." The ROD requires excavation of exceedances based on a point-by-point comparison with the RGs. This approach is consistent with past practice and with USEPA national guidance. Please revise the approach to require excavation of any exceedances based on a point-by-point comparison with the ROD RGs, excluding background and naturally occurring material.

9. **Section 3.3 and 4.3, Remediation Goals for soil and buildings, respectively:** These sections list the current ROD RGs. The HPNS's Five-Year Review occurring in 2018 is evaluating whether the current selected remedies, including these ROD RGs, are still protective and whether any changes are necessary to ensure continued protectiveness. Based on national practices directed by EPA headquarters, EPA expects this process to use the most current version of the EPA Preliminary Remediation Goal (PRG) Calculator and Building PRG Calculator to assess the ROD radiological RGs. The Work Plan should use only those cleanup goals confirmed through this analysis to be protective.

10. **Section 3.3.1, Investigation Levels, Table 3-6, Soil Survey Measurement Investigation Levels:** This section indicates that Investigation Levels are not applicable to the gamma scan surveys for Cesium (Cs)-137, and the footnote states that Cs-137 cannot be detected with the proposed gamma detector/gamma scan survey method at the RG of 0.113 pCi/g. Please describe how Cs-137 will be investigated in a manner that is compliant with a Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) survey design for which gamma scanning of 100% of the land area is completed with a detector capable of achieving the project-required detection limit and data quality objectives for the project. If the investigation will use alternative gamma measurement detectors with a better sensitivity that will allow Cs-137 to be identified at the RG above background (e.g. lanthanum bromide detector), then please revise the Work Plan to propose such a radioanalytical detection system. Alternatively, please revise the Work Plan to list the gamma scan survey achievable detection limit for Cs-137 and discuss how the survey(s) and sample collection will meet the data quality objectives for demonstrating that the survey unit is compliant with the ROD RG for Cs-137.

11. **Section 3.3.1, Investigation Levels:** The proposed investigation levels are inconsistent with the methodology proposed for the gamma scan surveys. Section 3.3.1, Investigation Levels,

states gamma scan surveys will be performed using detector systems equipped with gamma spectroscopy to provide real-time radionuclide-specific measurements, and the spectra will be evaluated using regions of interest peak identification tools for the ROCs that correspond to gamma rays at 186 kiloelectron volt (keV) for Ra-226, 609 keV for Bismuth-214 (bi-214), and 662 keV for Cs-137. However, the text does not state how the gamma scan can achieve sufficient detection limits for Ra-226 using the Ra-226 energy line at 186 keV due to the low efficiency at this energy or the Bi-214 609 keV line without a 21-day ingrowth period, especially when the investigation level is the same as the RG of 1 picoCurie per gram (pCi/g) above background. Additionally, Table 3-6 contains a footnote that states the gamma scan cannot achieve the detection limit necessary to detect Cs-137 at the RG of 0.113 pCi/g above background; yet the preceding text states that the gamma scan will be used to flag locations where Cs-137 exceeds the investigation level, defined in Table 3-6 as 100% of the RG, or 0.113 pCi/g above background. Please revise the Work Plan to address these concerns.

12. **Section 3.4.1, Number of Samples:** Although under some circumstances, 18 samples per survey unit could be acceptable as a default starting point before sampling results are available, once these results are available, then the number of samples for subsequent survey units should be based on calculations using variability found in actual data. For example, EPA's statistician used background data the Navy had previously collected from five reference areas and calculated that 25 samples per survey units would be needed to achieve your proposed 99% confidence level if soil from TUs/SUs are compared to reference background areas using a Wilcoxon Rank Sum (WRS) Test. EPA recommended starting with this default number of samples. Once new data are collected, they can also be used to recalculate the appropriate number of samples depending on the statistical tests which will be utilized to establish compliance. The new number of samples could be higher or lower than previously used.

    Note that the variance from site investigative samples may be larger than the variance based on reference background samples, therefore the variance from samples collected in investigative survey units should be used to calculate the number of samples that should be collected in other investigative survey units. Also, variance should be determined using the same radioanalytical method as that which will be used for additional data collection. For instance, the variance for gamma spectrometry laboratory data should be used to determine the number of samples that are required for survey units where gamma spectrometry laboratory analysis will be conducted.

    This section contains an inconsistent sampling scheme and does not comply with the requirements established in the Work Plan for number of samples required for each survey unit, as follows:

    a. It appears that the Work Plan does not provide the basis for the number of samples planned to be collected from TUs/SUs. The Navy previously issued a February 2018 Draft *Work Plan, Radiological Survey and Sampling*, which calculated the number of samples that would be collected from each SU using MARSSIM equation 5-1 for the WRS test. The Work Plan should use either the MARSSIM approach or other statistically based criteria for selecting the number of samples that will be collected from

each SU so that conclusions based on evaluation of the SU data can be defined by a statistical level of confidence and as such, are usable for decision-making. Please revise the Work Plan to include this information.

b. This section specifies the collection of twenty-five subsurface samples from each RBA location and twenty-five surface soil samples from the off-site (RBA-Bayview) location, but only requires five surface samples be collected from each of the on-site RBAs. The text does not state how or why it is appropriate to collect only five surface samples from each of the on-site RBAs when twenty-five samples will be collected from the surface of the off-site location, and twenty-five samples will be collected from each of the RBA subsurface areas. For the Bayview park off-site location, an important reason for sampling at this site is to get an indication of potential Cs-137 levels from fall-out, and to provide data that provides meaningful comparisons to on-site reference area data. Since on-site data will be collected from the surface and subsurface, the Work Plan should specify that both surface and subsurface data be collected from the off-site Bayview park location to provide a more complete and thorough evaluation of Cs-137 deposition and background levels in the San Francisco Bay area/the Hunter's Point Naval Shipyard. EPA understands that using a drill rig may present practical challenges to obtaining subsurface samples at the Bayview park location; therefore, the depth of subsurface samples collected will be based on the depth to which a hand auger can be used to collect the soil at the Bayview park. EPA appreciates the Navy's commitment to consult with a USGS Cs-137 expert in this process and in the field during sample collection. Please include this in the next version of the Work Plan and provide any comments from that expert in the eventual report that will be prepared about the sampling results.

c. The fifth bullet indicates that the total number of samples to be collected for surface soils in the on-site RBAs is twenty-five, but the text states that five samples from each of the four on-site RBAs will be collected, which is only twenty samples, not twenty-five. The text in this section and the bulleted information should be revised to provide a consistent number of samples.

d. Appendix A, Section 4.1.2, states that based on the statistical evaluations, the RBA report will include recommendations for combining similar data sets, recommendations for selecting values or data sets representing background in soil, and conditions identifying situations when specific values or data sets may not be appropriate. Since statistical testing will be completed to determine if each of the RBA data sets are sufficiently comparable to combine the data, please revise the Work Plan to discuss how/why five data points is sufficient for identifying a population that can reliably be compared to another five–point data set to determine if the difference is statistically significant or not.

Please revise the Work Plan to address these concerns.

13. **Section 3.4.4, Phase 1 Trench Unit Design:** The EPA, DTSC, and CDPH have prioritized trench units (TUs) for excavation using criteria listed in the EPA March 2018 Comments, e.g., Historical documentation of specific potential upstream sources, signs of potential falsification found in data evaluation, signs of data quality problems found in data evaluation,

allegations from former workers, and regulators' independent field testing. More details about these criteria are in the March 2018 EPA comments. In addition, EPA previously made comments to the Navy about the categories of concern in a letter to the Navy on February 27, 2017. The regulators' prioritization is partially consistent with the Phase I Soil Trench Units identified in the Navy's draft Work Plan. We concur with Phase I TUs 69, 76, 78, 99, 101, 103, 104,107, 108, 109, and 124. However, four of our highest priority TUs (TUs 97, 98, 115, and 121) are not included. These four TUs should be substituted for four of the 10 other TUs (i.e., those not listed above) that were identified as Phase I Soil TUs. Please make this change. The remaining soil TU's should be determined based on criteria such as those listed above, consistent with our March 2018 Comments and February 2017 letter that listed indicators of the highest likelihood of contamination. Choosing to prioritize a particular TU for logistical convenience due to TU's being adjacent is not an acceptable justification without independent evidence that this TU is among 33% of trench units most likely to have contamination based on the information we have to date.

14. **Section 3.4.6, Phase 1 Survey Unit Design, and Section 3.4.7, Phase 2 Survey Unit Design:** For the Soil Survey Units in former Building Sites, the same relevant comments already made on Sections 3.4.4 and 3.4.5 about trench units also apply to building site survey units.

15. **Section 4, Building Investigation Design and Implementation:** This section does not provide sufficient information to conduct a full evaluation of the sufficiency of the buildings investigation. The Navy's buildings data evaluation found significant enough extent of unreliable data that the Navy decided that none of the previous data could be used. Therefore, the presumption is that all previous work should be redone as a completely new investigation. Therefore, all specific details of a new building investigation/SAP should be provided in the Work Plan to adequately document the requirements of such an investigation. Please revise the Work Plan to specify a level of detail at least as thorough as typically done previously in Task Specific Plans for these buildings, as follows:

    a. Brief history of CSM along with a description of how survey units were identified and classified based on the CSM for each building, along with figures depicting the survey units and classifications, and sample locations.

    b. Complete listing of Data Quality Objectives (DQOs) for each Parcel, Survey Unit for land areas and for buildings

    c. All MARSSIM Final Status Survey (FSS) design parameters, including the identification of the survey unit classifications and sizes, and number of samples required to be collected for the WRS test, and all the associated calculation inputs, including the Lower Bound of the Gray Region, standard deviation of previously collected data, relative shift, confidence level selected, etc. This information should also include the identification of investigation levels for all radiological survey types, elevated measurement comparison calculations, or any other inputs and decision rules associated with the FSS design. In addition, when multiple radionuclides may be present, the Work Plan should include the identification of the

survey release limit and investigation level based on the sum of fractions and unity rule for all ROCs

d. Description of the Investigation Levels or other triggers that will be used in Gamma Scan Surveys that would require a biased sample to be collected

e. Listing of the specific radiological instrumentation that will be used for each scan and static survey, exposure rate measurements, and laboratory measurements with the associated achievable MDC, required scan rates, count times (statics), minimum detectable count rate (MDCR) for surveys; smear/wipe sample instrument MDCs, and laboratory instrument MDCs. MDCs should be 10% of the Remedial Goals for all ROCs

f. Inclusion of all the technical, as well as QC requirements for sample collection and analysis, data validation, assessment and reporting, along with copies of standard operating procedures (SOPs) for all of these processes. The technical information should include the method number, calibration information and quantitation parameters for scans, wipes, and static measurements. The QC information should include daily/weekly efficiency, energy and background checks as applicable; and results for duplicates, blanks, Laboratory Control Samples (LCS) samples (laboratory analysis), or matrix spikes and tracer recovery (only for destructive laboratory analysis) for each analysis type and instrument.

g. Copies of field and laboratory radioanalytical methods/Standard Operating Procedures (SOPs). SOPs should include the sample/aliquot size and count times needed to achieve the project-required detection limits at 10% of the RG, the error bars associated with the quantitation of all radionuclides, the nuclide library that will be used to identify the ROCs in the analysis, the data reduction and reporting procedures, and all instructions required to complete the analysis.

h. Reference to the appropriate Quality Assurance Project Plan (QAPP)/Master SAP which define all technical and quality parameters for data collection.

One possible approach the Work Plan could choose is to incorporate by reference some portions of the original Task Specific Plans for individual Buildings that are still relevant today, e.g. building description, building history, locations of survey units, extent of testing in categories of these survey units, etc. However, some other aspects of previous Task Specific Plans may need new scrutiny and potential modification in light of remediation that has already occurred, updated CSM information, new questions about reliability of prior work by Tetra Tech EC Inc., or other newly identified information.

Please revise the Work Plan to address the above specific issues for the buildings investigation.

16. **Section 4, Building Investigation Design and Implementation:** The Work Plan appears to depart from the previous practice of using the MARSSIM approach for identifying the

parameters of a FSS in order to demonstrate that a survey unit has met the release criteria. The parameters defined by the MARSSIM approach include the survey unit class and size, and include calculations for determining the number of samples that would need to be collected in each survey unit to meet the assumptions of the WRS statistical test with a specified level of confidence. The WRS test uses hypothesis testing to identify if the median of the site data is statistically the same or different than the median of the background data and as such provides a comparison of populations. This approach is well-established and accepted among many agencies for demonstrating that a survey unit has met the release criteria (derived concentration guidelines level [DCGL]) as determined by pathway modeling and exposure assessment.

However, the EPA regulates cleanups in accordance with the CERCLA statutes which require demonstrating that regulatory standards and/or risk-based target cleanup levels for hazardous substances will not exceed a specified limit, or pose an Excess Lifetime Cancer Risk to a reasonably maximally exposed (RME) individual that exceeds the CERCLA risk range of $1x10^{-4}$ to $1x10^{-6}$. Therefore, EPA has Superfund national guidance that recommends a more protective approach than MARSSIM in applying a point-by-point comparison between the investigative sample results and the RGs and which requires every exceedance of the RGs to be remediated. The more protective point-by-point approach has been used at the HPNS and most EPA Superfund sites nationwide for many years for both chemical and radiological cleanups. This approach must be included in this Work Plan. Even though this approach is more protective than what MARSSIM prescribes, the Work Plan should still use the MARSSIM approach to design the parameters of the FSS, as it has for many years, for consistency and defensibility of the results. Please revise the Work Plan to use the MARSSIM approach to design the parameters of the FSS and to require a point-by-point comparison between investigative sample results and the RGs, with remediation of areas where sample results exceed the RGs.

17. **Section 4, Buildings:** The number of samples determined to be required for building survey units should be calculated using the MARSSIM approach for the design of an FSS, and should be based on parameters obtained from collection of site samples of the same media and survey or lab instrument. These parameters include an estimate of residual radionuclide concentrations and the variance (σ) of results within a given survey unit or units. The value of σ may be obtained from earlier surveys, limited preliminary measurements, or a reasonable estimate. The estimate of σ includes both the real spatial variability in the quantity being measured and the measurement method uncertainty of the measurement method. Therefore, the initial number of samples may be based on information from previously collected data or may be estimated; however as newly collected data is obtained under the Work Plan, the variance used to determine the appropriate number of samples needed to meet the assumptions of the WRS test should be updated based on the variance from the new data. In addition, since the variance is a measure of spatial variability and the measurement method uncertainty, it is important that the variance from the same radioanalytical technique be used to estimate the number of samples being collected for the same analysis type. For example, the variance from newly generated gamma static surveys should be used to calculate the number of static measurements required in other survey units where static measurements are will be used for the FSS data collection.

Likewise, the variance from laboratory analysis of survey unit samples should be used to calculate the required number of samples needed to be collected in other survey units where the samples collected for the FSS will be analyzed by the same method in the laboratory. If the variance from newly collected data is smaller than that obtained from historical data or assumptions made about the population, then fewer samples may be needed for sample collection in other survey units. Finally, the variance from scan, static, smear, or sample analyses in the laboratory can only be used for sample number calculations of the same media type. Therefore, the variance obtained from gamma static surveys on land areas should not be used to calculate the required number of samples that will need to be collected in buildings. Currently, the Work Plan does not discuss the specifics of what variance will be used to calculate the required number of samples, or how newly collected data will be used to update the variance and the required number of samples in the FSS for on-going survey unit investigations. Please revise the Work Plan to describe in detail how the required number of samples will be determined for building survey units.

18. **Section 4, Building Investigation Design and Implementation:** In addition to the aforementioned deficiencies in the Work Plan Buildings investigation documentation, the following additional concerns require additional discussion, as follows:

    a. Section 4.1 (Data Quality Objectives) Step 5 – Develop Decision Rules states "If the investigation results demonstrate that site conditions are not compliant with the Parcel G RAO, then the data will be evaluated to determine whether site conditions are protective of human health using USEPA's current guidance on Radiation Risk Assessment at CERCLA Sites (USEPA, 2014)." However, EPA Directive 9200.4-40 was issued as guidance only and is therefore not a regulatory requirement, nor does it satisfy the ROD-established cleanup level for the Hunter's Point Naval Shipyard site in accordance with the CERCLA process as promulgated in 40 CFR §300.430. Please revise the Work Plan to state that only areas that demonstrate compliance with the Parcel G ROD requirements and are within the CERCLA risk range using the most recent version of the EPA PRG Calculator for radionuclides will be eligible for Regulatory Approval for release.

    b. The Work Plan does not explain why some buildings or portions of buildings will receive surveys and others will not. The Historical Radiological Assessment (HRA) Volume II should be used to summarize information about all buildings within Parcel G to provide justification for which buildings/areas will be surveyed. In addition, the justification should also include documentation from the data evaluation forms and conclusions regarding allegations of misconduct and fraud by the previous contractor, as well as Regulatory Agency input to this analysis.

    c. The text does not explain why MARSSIM Class 2 areas were not proposed around Class I areas when the entire building will not be surveyed.

    d. The Work Plan does not provide justification for selection of the area in Building 401 where background data will be collected.

e. The Work Plan does not discuss how the number of static measurements for each survey unit was calculated.

f. The Work Plan does not state if additional wipe samples may be sent to the laboratory for destructive analysis for speciation to determine which radionuclide is contributing to the radiation if release limits are exceeded for either gross alpha or gross beta.

g. The Work Plan includes a listing of the investigation levels but does not specify whether exceedance of the investigation levels will result in the collection of bias samples or static measurements in buildings.

h. The Work Plan does not specify collecting data from locations where measurements and/or sampling may be necessary due to use of equipment, areas where potential cross-contamination may have occurred, or where waste disposal practices may have resulted in contamination in sinks, or drains. Examples include items of equipment and furnishings, building fixtures, drains, ducts, and piping. Many of these items or locations have both internal and external surfaces with potential residual radioactive material which should be surveyed for removable and fixed contamination.

Please revise the Work Plan to address these concerns.

19. **Section 5.4 NORM Background Evaluation:** The proposed approach for performing a Naturally Occurring Radioactive Material (NORM) evaluation for site samples is insufficient for ensuring a complete and defensible analysis. The Executive Summary discussion of Data Evaluation and Reporting states "individual samples with gamma spectroscopy concentrations for Radium-226 (Ra-226) greater than the RG will be analyzed for Uranium-238 (U-238) and Ra-226 using comparable analytical methods. For that specific sample, the U-238 result will be used as a more representative estimate of the background value for Ra-226, and the Ra-226 concentration will be compared to the RG for Ra-226 using the revised background value." Per previous EPA comments, a sample with elevated Ra-226 above the RG should be analyzed for all uranium and thorium isotopes by alpha spectroscopy, and should be compared to data obtained in the gamma spectrometry analysis for all the radionuclides listed in the Appendix A, Table 3-6, Analytical Sample Summary. This information is required due to the following reasons:

a. U-238 results often have a large error bar/uncertainty associated with the result; therefore, analysis of other radionuclides in the U-238 decay series should be performed to confirm the accuracy of the U-238 result.

b. The alpha spectroscopy analysis for U-238 will also provide results for U-235 and U-234. All of the uranium isotopes reportable by alpha spectroscopy, including U-238, U-235, and U-234 should be reported in order to evaluate if the three uranium isotopes ratios indicate the uranium is present in natural abundance with uranium-238 at 99.2739–99.2752%, uranium-235 at 0.7198–0.7202%, and uranium-234 at 0.0050–0.0059%.

c. Alpha spectroscopy analysis of thorium isotopes (Th-230 and Th-234) is requested to confirm the Uranium-238 result since Th-234 is the first daughter product of U-238. In addition, Th-230 is the immediate precursor to Ra-226 in this series; therefore, analysis of this isotope will help confirm whether the U-238 decay series is in equilibrium.

d. Gamma spectrometry analysis for Bismuth and Lead isotopes that are part of the Thorium and Uranium decay series. Potassium-40 (K-40) will provide further evidence of whether the ROCs detected in the analysis are from naturally occurring background or represent contamination.

Please revise the Work Plan to require all samples with elevated Ra-226 results to be analyzed for all Uranium and Thorium decay series isotopes by alpha and gamma spectroscopy to provide sufficient documentary evidence regarding the NORM evaluation.

20. **Section 5.6, Reporting, Page 5-7:** The text indicates that where a TU/SU exceeds the Parcel G ROD RAOs, the Removal Site Evaluation Report will include recommendations and options for further action, including the possibility of revising the Parcel G ROD to demonstrate the unit has met compliance criteria. However, the current compliance criteria are the Parcel G ROD RGs. Unless the Navy performs an analysis that demonstrates that the current RGs are no longer protective (for instance, by evaluating the RGs using the most current EPA PRG calculators), an amendment to the Parcel G ROD would be unnecessary. Therefore, please revise Section 5.6 to remove reference to revision of the Parcel G ROD as a potential solution to demonstrating a TU/SU meets the release criteria in the Work Plan.

21. **The Appendix A, Soil Reference Background Area**: This section does not reference a Quality Assurance Project Plan (QAPP) or a task-specific work plan/sampling and analysis plan (TSP/SAP) which specifies the details of all quality and procedural requirements for this data collection project. Please revise Appendix A to include this information.

22. **Appendix A, Soil Reference Background Area**: It is unclear whether the proposed background locations are suitable for collection of background samples because the Work Plan does not provide details about these locations. For example, it is unclear if there were any previous excavations (e.g., exploratory excavations, remedial excavations, fuel line removals, or sanitary sewer/storm drain removal excavations) in these areas. If any of these areas have previously been excavated, then it would be unsuitable for use as a reference background area (RBA). In addition, the location proposed in Parcel D-2 is near the foot of a steep slope where erosion and run-off may have concentrated radionuclides found in atmospheric fallout like Cs-137; if this is the case, this location is unsuitable as a background location. Further, the location proposed for Parcel UC-2 is near or at the bottom of a hillside, where runoff may also have concentrated Cs-137 and be unsuitable for use as a background site. Although the text describes these areas as "non-impacted," a detailed justification for each proposed background area should be provided. Please revise the text to include a detailed justification for each proposed background location and exclude any locations where erosion and runoff may have concentrated radionuclides found in atmospheric fallout.

23. **All sections:** EPA appreciates the multiple technical meetings with the Navy to discuss these comments and the verbal commitments from the Navy to revise the Work Plan to address many of these comments. We look forward to seeing the revised Workplan that incorporates these changes. EPA is making every effort to include in our formal comments every concern that we may have. If significant new information comes to light, including related to public comments, or significant new insights result from further evaluation, EPA may supplement these comments later.

24. All Sections: Attached is a statistical review of the Work Plan that applies to all sections of the Workplan.

SPECIFIC COMMENTS

1. **Executive Summary:** The next draft of the Work Plan will receive a great deal of attention from the public. Laypeople reading it would benefit from a summary that is more understandable to a general audience, e.g. similar to the fact sheet that the Navy already distributed June 2018 to accompany its draft Work Plan. EPA recommends that the Navy update its fact sheet to reflect the next draft version of the Work Plan, distribute that updated fact sheet to the public, and insert the updated fact sheet into the beginning of the next draft before the Executive Summary.

2. **Executive Summary, p. iii, and Table 2-1, Conceptual Site Model – Uncertainties, Page 2-5:** The Executive Summary references "Allegations of previous sample collection fraud, . . . " and the Table references "Potential for data manipulation or falsification." Yet some instances of these practices have been confirmed by the 2014 Tetra Tech EC Inc. Internal Investigation, the 2016 Nuclear Regulatory Commission concluded enforcement action, and the 2018 Department of Justice concluded criminal cases that sent two people to prison. Please add language to state that some fraud, manipulation, falsification, etc. have been confirmed.

3. **Section 1, Introduction:** For context to the reader, please clarify that other future work plans will address other aspects of the site where Tetra Tech EC Inc. has previously performed radiological work. For example, EPA commented in March 2018, "Tetra Tech EC Inc. also conducted radiological cleanup work in ship berths. The Navy should also address potential contamination in this and any other category of past radiological work by Tetra Tech EC Inc. at the HPNS." Please insert language into the Work Plan to convey this larger context into the introduction.

4. **Table 2-1, Conceptual Site Model – Uncertainties, Page 2-5:** The Uncertainties section of Table 2-1 states storm drains and sewer lines, including one foot of soil surrounding the pipes were removed to within 10 feet of all buildings, and impacted buildings had the remaining lines removed during surveys of the buildings. Non-impacted buildings are stated to have had surveys performed at ends of pipes and were capped. However, review of the Parcel G Data Evaluation Forms identified several instances of pipes being found in areas where they were thought to have previously been removed.

Please revise the uncertainty discussion in the Table 2.1 Conceptual Site Model to list this additional uncertainty.

5. **Table 2-1, Conceptual Site Model – Uncertainties, Page 2-5:** The "Uncertainties" section states that an example of a factor that results in a lower potential for radiological contamination is power washing. However, the "Potential Migration Pathways" section on Page 2-4 of the same table lists power washing also. Furthermore, the Navy's 2008 Technical Memo, Section 3b. Conceptual Site Model, states that power washing increases the potential for cracks in piping that could increase seepage of radiological material into the surrounding soil.[7] These appear to contradict. Please remove power washing from the list of factors that could lower the potential for radiological contamination.

6. **Section 3.3 and 4.3, Remediation Goals for soil and buildings, respectively:** Please revise the Work Plan to explain how compliance with RGs will be evaluated when more than one ROC is identified. Cleanup goals should include an analysis of the sum of fractions and the unity rule to ensure total risk to the Reasonably Maximally Exposed (RME) individual posed by multiple ROCs in soil or buildings does not exceed the CERCLA risk range of 1 X 10^{-4} to 1 X 10^{-6}. Please note that "Consistent with existing Agency guidance for the CERCLA remedial program, . . . EPA generally uses 1 x 10^{-4} in making risk management decisions."[8]

7. **Section 3.3.1, Investigation Levels:** This section defines investigation levels as media-specific, radionuclide-specific concentrations, or activity levels based on the remediation goals (RGs) that trigger a response, such as further investigation, if the investigation level is exceeded. The text also states that investigation levels are established for each instrument and vary with SU classification and measurement type. It is unclear, however, why the investigation levels may vary by survey unit. Please remove text that indicates that the investigation levels would vary by survey unit.

8. **Section 3.4, Radiological Investigation Design:** The Parcel G Work Plan does not consider the need to investigate contamination associated with radiological objects containing Strontium-90. A gamma scan survey can be used to detect the bremsstrahlung radiation caused by Sr-90, but the text does not discuss collection of this data. Please revise the Parcel G Work Plan to discuss how the potential presence of Sr-90 in soil will be assessed.

9. **Section 3.5.2.2, Site Preparation, Page 3-13:** The second to the last bullet point states that after removal of the durable cover, "an additional 1 foot of durable cover buffer beyond the former excavation surface boundary will be removed," but the Navy response to EPA Specific Comment 16 states that "anything removed will be surveyed." Please revise the text to discuss whether excavation of this additional foot of soil is sufficient to account for regrading and clarify if this soil will be scanned and sampled or sorted.

[7] *Navy Memorandum for the Record: Conceptual Site Model for the Removal of the Sanitary and Storm Sewers at Hunters Point Shipyard*, December 17, 2008
[8] OSWER Directive 9200.4-40, EPA 540-R-012-13, May 2014, Q34, p. 27.

10. **Section 3.6.4, Phase 2 Trench Unit Investigation:** Three samples should be collected at each core, including those less than 4 feet in depth below ground surface (bgs). Please revise this section to specify three samples will be collected for each core regardless of the depth of the core.

11. **Section 3.6.3.1, Automated Soil Sorting System Process, Page 3-15**: It is unclear if a single sample of the diverted soil material will be sufficient to characterize this material, particularly if there is a large volume of diverted soil. Since soil can be diverted for reasons other than radiological alarms (e.g., low mass on the conveyor belt), it is important to collect sufficient samples to characterize this soil. Please revise the Work Plan to propose a volume-based sampling protocol with a one-sample minimum to characterize diverted soil.

12. **Section 3.6.3.1, Automated Soil Sorting System Process:** This section provides a description of one alternative for gamma scans to screen soils from TUs/SUs at Parcel G, but clarification is needed:

a. Please include a description of the detectors that will be used or the detection limits of those detectors.

b. The text states that soils will be sorted based on radiological properties. Please provide specifics about which radiological properties will be monitored and used for segregation. Please explain if the alarm will be set to an investigation level or if it will be set at multiple levels such that alarms occur when one of the ROC RGs or investigation level is exceeded.

c. This section indicates the details of such an operation are included in the Soil Sorting Operations Plan, but this Plan was not included in the Work Plan. If this option is chosen, the Soil Sorting Operations Plan should be submitted for Regulatory Agency review and approval before soil sorting is implemented.

Please revise the Work Plan to address these concerns.

13. **Section 3.6.4, Phase 2 Trench Unit Investigation, Page 3-17:** It may not be possible to collect cored samples to 6 inches below the depth of the original excavation if gravel was used to bridge the water table when the original excavation when backfilling occurred. Many of the open trenches in Parcel G contained groundwater because the water table is relatively shallow, so it is likely that gravel may have been used as backfill in some or all of these trenches. Trenches where gravel was or may have been used to bridge the water table should be identified so that an alternative sampling method (e.g., potholing) can be used. Please identify trenches where gravel was or may have been used to bridge the water table and propose an alternative sampling method to obtain samples from 6 inches below the depth of the original excavation.

14. **Section 3.6.4.1, Subsurface Soil Sample Collection, Page 3-18:** The text indicates that "use of a 3-inch-internal-diameter sampler may be required" in order to obtain sufficient

sample volume for analysis, but it is unclear why drilling a second borehole adjacent to the first is not included as a potential method to collect sufficient soil. If the soil is sandy, it may not be retained in a 3-inch sampler, but may be retained in a smaller diameter sampler with a bottom basket. Please propose multiple potential methods for collection of sufficient sample volume.

15. **Section 3.6.7.2, Decontamination and Release of Equipment and Tools, Page 3-21:** The text discusses possible decontamination of equipment and tools at the completion of fieldwork, but this should not be optional because there could be chemical contamination in addition to radiological contamination. In addition, decontamination of equipment and tools is necessary between sampling locations (e.g., shovels, trowels, mixing bowls, coring equipment). Please revise the Work Plan to provide a more complete decontamination plan and to require decontamination of all equipment and tools before they are removed from the site.

16. **Section 5.5, Reference Background Area Soil Data, Page 5-6:** The text states that RBA data sets will be compared to each TU/SU data to demonstrate the RBA data set for soil is representative of soil in each TU/SU by comparing the median of the two data sets to determine if there is a statistical difference in the medians. However, the text does not state how it will be determined that the soil sample(s) collected from the TU or SU used for this comparison will represent only background and not site contamination. Further, it is unclear why the Work Plan proposes to compare the medians of data populations between background soil and investigation unit soil rather than to perform the evaluation recommended by the EPA. This evaluation includes analyzing the soil for the primordial naturally occurring parent and daughter radionuclides to determine if they are in secular equilibrium to identify whether the radionuclide ratios indicates the soil represents background. Please revise the Work Plan to require evaluation of secular equilibrium before any statistical comparisons are conducted.

17. **Appendix A, Section 2.0, Purpose and Objectives, Step 2 - Identify the Objective, Page 2-1:** The text does not appear to distinguish between potential contamination and background levels. Step 2 states that the background study is being conducted to "establish representative background data sets for soil ROCs, NORM radionuclides, and fallout ROCs for comparison and evaluation of soil data collected from the HPNS." This statement seems to imply that soil ROCs may be present in background that are not present due to Naturally Occurring Radioactive Material (NORM) or from fallout associated with nuclear tests or reactor accidents. There is a similar statement under Step 3 - Identify Inputs to the Objective. Please revise the text to clarify that only ROCs that are present due to NORM or fallout may be considered background.

18. **Appendix A, Section 2.0, Purpose and Objectives, Step 4-Define the Study Boundaries, Page 2-1:** Step 4 proposes an inconsistent sampling strategy. This section states that in Parcels B, C, D-1, and D-2, reference background surface soil samples will be collected from 0 to 6 inches below ground surface bgs, and subsurface soil samples will be collected from 1- to 2-foot intervals to a depth of up to 10 feet bgs. However, at the off-base location, surface soil samples will be collected from 0 to 6 inches bgs and

subsurface samples to a depth of 10 feet bgs are not proposed. It is unclear why samples collected from on-base background locations will be obtained from the subsurface in 1- to 2- foot internals to a depth of up to 10 feet bgs, but off-site background samples will only be collected from 0 to 6 inches. Collecting subsurface samples from the off-site location will provide valuable information about the depth of deposition and transport of radionuclides from fallout, as well as the potential differing distribution of NORM at depth. In addition, a lithological profile of off-site subsurface soil should be completed to provide additional support to any correlation drawn from soil profiles and NORM collected at the HPNS. Please revise the off-site sampling approach to include collection of subsurface samples.

19. **Appendix A, Section 2.0, Purpose and Objectives, Step 5 - Develop Decision Rules and Step 6 - Specify the Performance Criteria, Pages 2-1 and 2-1:** The performance criteria discussion states that the background data sets will be evaluated for suitability based on statistical tests, but prior to performing the statistical tests, an evaluation of whether the naturally occurring radionuclides that are also ROCs should be evaluated to determine if the U-238 parent and daughter radionuclides, and as applicable, Th-232 and daughter radionuclides are in secular equilibrium. This is necessary to ensure elevated ROCs that are present due to contamination are not eliminated as outliers. Please revise this discussion to address the need to evaluate whether the U-238 and Th-232 series radionuclides are in secular equilibrium before performing statistical tests to identify outliers or to derive population estimators for comparison to site data.

20. **Appendix A, Section 3.1.6, Field Instrumentation, Gamma Detectors, Page 3-5 and 3-6:** The text provides a list of two gamma survey instruments that will be used in the RBA but does not provide the detection limits for each instrument. Please revise Appendix A to include the efficiency and detection limits for the gamma survey instruments and the required instrument sensitivities that meet the data quality objectives for identifying radionuclides at background levels.

21. **Appendix A, Section 3.1.7, Laboratory Analysis, Pages 3-6 and 3-7:** Section 3.1.7 lists the radionuclides that will be analyzed but does not reference the QAPP that contains the QC requirements or detection limits for such analysis. Please revise Appendix A to include this information or reference the QAPP that includes this information.

22. **Appendix A, Section 3.2.4, Surface Soil Sampling Process, Pages 3-9 and 3-10 and Section 3.2.5.2, Subsurface Soil Sample Collection, Pages 3-11 and 3-12:** Please specify the required sampling volume and sample container in Section 3.2.4 and Section 3.2.5.2. Similarly, please specify the type of container that will be used to store soil intervals not designated for sampling (e.g., will core boxes or sealed jars be used?).

23. **Appendix A, Section 3.2.4, Surface Soil Sampling Process, Pages 3-9 and 3-10 and Section 3.2.5.2, Subsurface Soil Sample Collection, Pages 3-11 and 3-12:** Please provide decontamination procedures for drill rig tooling, hand tools, and bowls used for mixing should be specified in the text.

24. **Appendix A, Soil Reference Background Area Work Plan, Section 4.2 Analytical Data Evaluation:** The Work Plan in Appendix A should be revised to provide a more comprehensive strategy for selecting background values for comparison to site data and use in demonstrating compliance with the ROD RGs. For example, the strategy should consider the following inputs: the population distribution, characteristics (i.e. skewness) and variance for each background reference location or multiple locations; the frequency of detection; and site-specific factors (i.e. soil type, topography, depth, homogeneity or heterogeneity of the data set, or other). In addition, analysis of the background data set should include the appropriate statistical calculations or charts and graphs (such as quantile-quantile [Q-Q] Plots). The Work Plan should also describe how background data sets will be validated and at what frequency and should state that the complete data packages and data validation reports will be made available to the regulators for review prior to the selection of background values. Finally, one or more scoping and decision-making discussions between the regulators and the Navy should be conducted to select the most appropriate background values. Please provide a more comprehensive strategy for selecting background values that includes these issues. In addition, please revise Appendix A to specify that the full data packages, data summary tables, and data validation reports (from third-party data validators) will be given to the regulators for review.

25. **Appendix A, Section 4.2.2 Identify Outliers, Page 4-2:** This section states that background data values will be evaluated to determine if any are outside of the expected distribution using Dixon's and Rosner's statistical outlier tests, both of which assume the data are normally distributed. However, the previous Section 4.1.2, Outliers Test, states, "Because environmental data tend to be right-skewed, a test that relies on an assumption of a normal distribution may identify a relatively large number of mathematical outliers." Section 4.1.2 also states that outliers identified in statistical test will be reviewed to determine whether any suitable reasons (e.g., a potential analytical error) exist to exclude them from further calculations, and confirmed outliers will be removed from individual data sets. Therefore, please revise the Work Plan to specify that all background data sets should be evaluated using non-parametric statistical tests to evaluate population estimators (i.e., such as mean, standard deviation, and others) and potential outliers. Also, please ensure all naturally occurring radionuclides that are also ROCs undergo an evaluation to determine if the U-238 and Th-232 decay chains are in secular equilibrium prior to conducting any outlier evaluations to ensure ROCs that are present due to contamination are not eliminated.

26. **Appendix A, Section 4.3, Reporting, Page 4-4:** This section states that information from other San Francisco Bay Area radiological background studies may be referenced in the BRA report as appropriate. Please also revise the Work Plan to state how the Navy will determine if the other San Francisco background data sets are sufficiently comparable/representative of conditions/soils at the Hunters Point Shipyard.

ATTACHMENT

Statistical Review

leidos

DATE: August 2, 2018

TO: Felicia Barnett, Director SCMTSC, EPA WAM

FROM: Donna J. Getty, SERAS Statistician

THROUGH: Richard Leuser, SERAS Deputy Program Manager/Task Leader

SUBJECT: FINAL - STATISTICAL REVIEW OF UNITED STATES NAVY DRAFT PARCEL G REMOVAL SITE EVALUATION WORK PLAN, HUNTERS POINT NAVAL SHIPYARD SITE (HPNS), SERAS-106, WORK ORDER #83

INTRODUCTION

United States (US) Environmental Protection Agency (EPA) Region 9 personnel requested a statistical review of the sampling strategies outlined in the US Navy's *Draft Parcel G Removal Site Evaluation Work Plan Former Hunters Point Naval Shipyard, San Francisco, California* (June 2018) (Work Plan). The Naval Work Plan presents an overview of proposed methodologies and protocols for the investigation of radiological contamination in Parcel G and a proposed strategy for establishing background radiological levels, which is found in Appendix A of the document, *Draft Soil Reference Background Area Work Plan.* The proposed work is being conducted with the current knowledge of allegations that previous sampling and analysis efforts conducted by a Navy subcontractor may have included sample collection fraud, data manipulation, improper sample custody, and other issues that may indicate that radiological contamination is still present on Parcel G. US EPA Regulators have previously commented on an over-arching Draft Work Plan put forth by the Navy in February 2018 including the general methodologies which the Navy is proposing to use across all Parcels at the Hunters Point Naval Shipyard (HPNS) (Site). Some comments contained in this review refer back to these previous comments and in particular to the statistical review provided by Leidos on March 15, 2018.

GENERAL COMMENTS - PARCEL G

It does not appear that prior comments made in the Leidos statistical review regarding proposed sample sizes for background and site areas, statistical testing, confidence levels, interpretation of results, and application of *Multi-Agency Radiation Survey and Site Investigation Manual* (MARSSIM, 2002) methodologies were considered in this version of the Draft Work Plan. Methodologies including comparison of individual measurements to remediation goals (RGs) (based on background levels), without including inferential statistics to support decisions regarding compliance, are still the proposed sampling strategy. Sample sizes are still not statistically based and do not reflect the estimated variability from previous background sampling efforts. This is true for the on-site sampling strategy and the background study design. The on-site trench unit (TU)/surface soil survey unit (SU) assessment does not include statistically or probability-based testing. A clear statement of how statistical confidence will be included in the assessment of Parcel G is lacking from this Draft Work Plan.

Missing from the Work Plan is a clear statement as to the data quality objective that the Navy is planning to achieve. NUREG-1505 discusses the difference between two potential scenarios that can be selected for the assessment and closure of radiological sites; they are referred to as Scenario A and Scenario B. The choice of scenarios is based on whether the Navy is assuming Parcel G still contains levels of radionuclides of concern (ROCs) that are not in compliance and collecting data to prove otherwise (Scenario A) or the Navy is assuming Parcel G soils are in compliance and collecting data to prove otherwise (Scenario B). There is a critical distinction between the two Scenarios. The burden of statistical proof is always on proving the original assumption is incorrect. If the Navy opts for Scenario B, then the burden of proof rests on collecting a sufficient number of representative samples to refute the

assumption that Parcel G soils are in compliance. As put forward throughout this technical memorandum (TM) and as understood by this reviewer, the current Navy Draft Work Plan does not include a sufficient number of samples nor appropriate statistical testing to provide enough proof to refute the assumption that Parcel G soils are in compliance. A detailed explanation of additional concerns is provided below.

Parcel G: Establishing Compliance for Parcel G Soils

As stated in the Draft Work Plan (Section 3.1, page 3-1), the objective of the soil investigation is to determine whether concentrations of ROCs in soil are compliant with the Parcel G Record of Decision (ROD) by meeting the remedial action objective (RAO) for radiologically impacted soil. The RAO for Parcel G is to prevent exposure to ROCs in concentrations that exceed RGs for all complete exposure pathways. Logically, this can be achieved either by determining compliance for all of Parcel G by evaluating 100 percent (%) of the areas of concern located on Parcel G or through the selection of representative subsets of the areas of concern and applying statistical and probability theory to establish a level of confidence to the determination of compliance for the remainder of the Parcel.

The Navy's Draft Work Plan outlines a strategy which provides neither 100% evaluation of all areas of concern in accordance with the original process of excavating, sampling and scanning 100% of the TUs nor application of statistical/probability theory to a subset of areas to establish compliance for the entire Parcel. The proposed strategy includes performing an in-depth sampling/analytical and scanning evaluation of 21 of the TUs located on Parcel G and 14 of the 28 SUs, former building sites, located on Parcel G; this is identified as Phase 1 of the Navy's approach. Neither justification for how the Navy derived the number of TUs and SUs to be evaluated is provided, nor is the level of confidence that will be associated with determining that all of Parcel G soils comply with the RGs. As stated earlier, the original work conducted on Parcel G, the results that are in question, involved the excavation and modified *Multi-Agency Radiation Survey and Site Investigation Manual* (MARSSIM, 2002) Class 1 evaluation of 100% of the TUs and SUs. Because of the uncertainty associated with the historical data, the underlying questions include: 1) what sampling/scanning process needs to be followed to determine if the soils of Parcel G are compliant with the ROD specifications and 2) whether the project team wants to be able to associate a statistical confidence level with their assessment of compliance.

Statistical confidence can be achieved using statistical and/or probability analysis; however, statistical assumptions and conditions must be met prior to establishing that confidence. On-going discussions between the Navy and Regulators (Federal and State personnel) has resulted in an EPA proposed methodology of applying an "item sampling" statistical strategy. This will achieve a level of confidence that a stated percentage of the TUs and SUs are in compliance without re-excavating/re-sampling 100%, the level of effort that determined to be required to achieve the RAO in the historical work enacted by the Navy. These calculations were provided to the Navy with the following assumptions/conditions:

- The number of TUs/SUs, which would require a full MARSSIM Class 1 evaluation or otherwise agreed upon rigorous evaluation (referred to as Phase 1 in the Work Plan), was calculated to be able to demonstrate, with high probability (95% confidence) that a high percentage (95%) of the TUs/SUs were acceptable (concentrations below RGs).
- The calculated number was based on the assumption that historical information provides enough evidence to establish which buildings and TUs/SUs are at least two times more likely to be contaminated, and those are the TUs/SU that will be excavated, sampled and scanned.
- For each TU/SU evaluated by applying a rigorous Class 1 evaluation, a final binary answer/decision would be asserted: "Yes", the TU/SU is compliant, or "No", the TU/SU is not compliant. The conditions/tests for establishing compliance would be clearly stated so there would be no confusion in the final decision.
- The confidence statement being tested is: "Given the assumption that the 21 targeted TUs are 2 times more likely to be unacceptable than a randomly selected TU; I want 95% confidence that at least 95% of the TUs are acceptable."
- If any of the selected TUs fail compliance then:
    1. The confidence statement no longer applies; it must be assumed that at least a portion of the TUs are unacceptable and the population of Parcel G TUs fail compliance.

2. The assumptions made to determine the appropriate number of TUs to sample are no longer valid.
3. 100% of the TUs on Parcel G would have to undergo a Class 1 or similar investigation to assure compliance for Parcel G TUs.

A similar scenario would result for SUs that failed compliance. Based on the above assumptions and conditions, it was computed that if 21 of the 63 TUs were evaluated and demonstrated compliance, the Navy and EPA could be 95% confident that at least 95% of the 63 TUs are in compliance. Similarly for a population of 28 SUs, 15 should be evaluated and found in compliance.

This type of sampling is typically associated with batches of items, such as drums or wipe sampling for asbestos on surfaces, with clearly defined criteria for acceptance. There is some uncertainty associated with this statistical analysis given that the decision criteria for acceptance of the TUs and SUs will be derived from multiple lines of evidence instead of a single clear-cut measurement of a characteristic (i.e., flammability of waste in a drum, presence of asbestos on a wipe sample) as is typically associated with item sampling. Because of this unquantifiable uncertainty, the EPA regulators have requested that the remaining TUs and SUs still undergo assessment, which will allow investigation without excavation. This is to increase the overall confidence in the determination of compliance for Parcel G.

As stated above, the current Naval Draft Work Plan does propose evaluating 21 of the 63 TUs, which is in agreement with the proposed item sampling calculations; however, justification and associated levels of confidence are not identified. Without statistical justification, confidence in the final conclusion of compliance or non-compliance is uncertain. If application of the item sampling strategy is the justification for n=21, then the assumptions and conditions of this type of sampling must be met to establish a 95% confidence level that soil in a minimum of 95% of the TUs/SUs contain concentrations of ROCs below the RGs on Parcel G of the HPNS. In line with the EPA Regulators, Phase 2 of the Work Plan does call for a less rigorous evaluation of the remaining 42 TUs and 14 SUs; however details of the sampling strategies which will be employed to conduct the evaluations are still under discussion.

Parcel G: Establishing Compliance per TU and SU

The objective of this field sampling effort is to establish whether soils on Parcel G are compliant with the RAO and therefore below the RGs. As discussed above, it is uncertain how compliance for all of Parcel G will be ascertained based on the subsets of selected TUs (n=21) and SUs (n=14) for Phase 1, as there is no statistical confidence associated with selection of the size of the subsets in the Draft Work Plan for Parcel G.

Once an appropriate number of TUs/SUs for a modified Class 1/Phase 1 evaluation has been established, the next part of the assessment is the process for determining compliance for individual TUs/SUs. As presented in the Draft Work Plan, compliance will be determined by comparing each ROC measurement collected within a TU/SU with the corresponding RG. If any measurement exceeds the RG, compliance will not have been established for that TU/SU. This is contrary to methodologies in MARSSIM that recommend statistical testing and specified confidence levels to support compliance decisions. Table 2.2 of MARSSIM provides recommended conditions for demonstrating compliance: for Class 1 Survey Units a statistical test on sampling results, with elevated measurement comparison, and 100% scanning is recommended. For sites where ROCs are present in background, MARSSIM recommends the Wilcoxon-Rank-Sum (WRS) Test for the statistical test, although alternative statistical testing is permitted. MARSSIM recommendations do not include a comparison of individual measurements to the RG, although these comparisons could be made in addition to the statistical testing. Page 2-8 of MARSSIM states: "*A decision is made, in coordination with the responsible regulatory agency, based on the conclusions drawn from the assessment process. The ultimate objective is to make technically defensible decisions* ***with a specified level of confidence.***"

The following text is taken from the March 2015 Leidos TM:

There is statistical justification for requiring the WRS test or another test, which accounts for variability and distributional characteristics of the sample data. Statistical tests such as the WRS test are hypothesis tests, which are based on statistical inference. Statistical inference permits one to generate a conclusion about population characteristics based on information provided by a sample collected from that population. It provides a means for comparing the characteristics of one population sample to another population sample. In other words, the conclusions

drawn from these tests can be applied to all of the un-sampled components of the population. There are also specified statistical levels of confidence associated with these tests.

The proposed comparison of individual measurements collected within a TU/SU to the RG, results in a decision of compliance/non-compliance, which is only applicable to those individual measurements and cannot be extrapolated to the soils in the TU, which were not sampled.

Further text from the March 2015 TM states:

It is possible for samples collected within an SU to exceed the release criterion, even if the conclusion based on the WRS test is that the SU meets the release criterion. Because of the possibility of the presence of a few elevated concentrations, MARSSIM does recognize the need to support release/remedial efforts by comparing elevated measurements to the release criterion. However, this is done in addition to the WRS test not instead of the WRS test. As stated earlier, results of those comparisons cannot be extrapolated to soils beyond where the discrete samples were collected with any statistical confidence.

Alternately, an upper tolerance limit (UTL) can be computed per TU/SU and compared to a specified action level, such as a release criterion; however, sample size calculations need to be based on the computation of these limits and not on the WRS Test.

Discussion: The methodology proposed by the Navy for assessment of Parcel G does not include a measure of statistical confidence. It is widely recognized (MARSSIM 2002; NUREG-1505 Rev. 1) that statistical confidence is a critical component in establishing a defensible decision that a radiological site is in compliance. It is highly recommended that the Navy incorporate statistical testing into their data quality objectives (DQOs) so that conclusions drawn from this sampling effort can be substantiated and defended, particularly given the significant scrutiny currently surrounding the HPNS Site assessment.

Sample Sizes per TU/SU

Section 3.4.1, page 3-4 states that the number of samples to be collected per TU/ SU will follow the previously established protocol of a minimum of 18 systematically located samples. A discussion of sample size was presented by Leidos on March 15, 2018 demonstrating that following the historical protocol of n=18 samples may not achieve the desired confidence level and associated power of 99% (α=0.01 and β=0.01). It needs to be re-iterated that confidence is obtained through the application of inferential statistics and not through the Navy's proposed methodologies for Phase 1; therefore it is unclear how confidence in the decision of compliance/non-compliance will be obtained from the Navy's strategy.

It is assumed by this reviewer that n=18 was selected because historical data was not available to provide representative estimates of expected variability, therefore default assumptions were made based on MARSSIM Chapter 5 for a WRS Test. These default assumptions result in the computation of a sample size of n=18 (Table 5.3 of MARSSIM) for both on-site and reference background areas (RBAs). Calculations presented in the Leidos March 2015 TM demonstrate that n=18 may be an insufficient number of samples for conducting a WRS Test with 99% confidence. In the TM, samples sizes were computed based on the standard deviations computed from the historical sampling results of the original RBA's. Standard deviations ranged from 0.27 to 0.47 picoCuries per gram (pCi/g) for Radium-226 and from 0.031 to 0.046 pCi/g for Cesium-137 in the historical RBAs. Expecting greater variability in the on-site data than in the RBA data, and to be conservative and protective of human health, the maximum standard deviations from the RBAs were used in calculating sample size for α=0.01, and β=0.01. Applying the MARSSIM methodology a sample size of n=25 was computed. If the WRS test will be applied to obtain statistical confidence in deciding compliance/non-compliance for individual TUs/SUs, sample size should be increased from 18 to 25, unless evidence can be provided to support n=18.

As discussed in the March 2015 TM:

Alternately, a UTL can be computed per SU and compared to a specified AL, such as a release criterion. Using a previously computed release criterion for Ra-226 at the HPNS site, 2.4 pCi/g, exploratory computations were

performed to determine the sample size required for the computation of a non-parametric UTL using Visual Sample Plan (VSP) software. For at least 95% confidence that 95% of the population of surface soil within an SU has Ra-226 measurements below the AL, 59 samples would need to be collected. The non-parametric UTL was chosen to parallel the non-parametric choice of the WRS test by MARSSIM.

Discussion: Regardless of the statistical test, which is selected for establishing compliance per TU/SU, the Work Plan should include the calculations and the inputs to the calculations for the sample size that will be used. If n=18 is the correct sample size, include the calculations and data inputs that support that value. If a statistical test will not be utilized, please provide a discussion in the Work Plan with an explanation as to how confidence in the decision of compliance/non-compliance can be quantified.

GENERAL COMMENTS- ESTABLISHING BACKGROUND – APPENDIX A

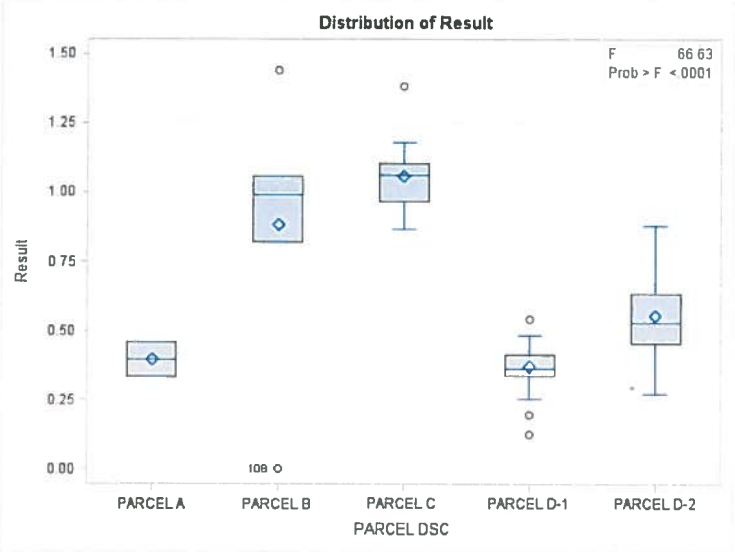
Five background reference areas have been selected for evaluation. Four areas have been selected on the HPNS which have been identified as non-impacted by the historical activities conducted on the Site (RBA-1 through RBA-4) and one off-site area, Bayview, has been selected for characterization of "ubiquitous fallout radionuclides". Surface and subsurface soil samples will be collected from the four on-site RBAs with only surface soil samples collected from the Bayview Area. Comments presented in the Leidos March 2015 TM regarding the sampling strategy for the proposed RBAs have not been addressed in the current Background Work Plan and will be revisited in this review.

Establishment of representative background ROC levels is critical for determining the RGs for the HPNS Site assessment. For example, the RG for Radium-226 is stipulated to be 1 pCi/g above background and therefore is dependent on an accurate and representative background evaluation. A sufficient number of samples need to be collected from each RBA per strata (surface and subsurface) to ensure the calculation of defensible background for the ROCs. Presently, it is unknown whether surface and subsurface ROC concentrations in the RBAs belong to the same statistical population. Until formal statistical testing is conducted to verify they are from the same population they should be examined independently when computing appropriate sample sizes. It is important to acquire enough information during the current sampling effort to establish defensible and representative background without returning for an additional round of sampling.

Sample Sizes per RBA

As per the discussions of sample size for the on-site TUs and SUs, the appropriate sample sizes should be determined based on estimated variability from the historical sampling efforts. If the WRS Test will be run, MARSSIM states that sample sizes for the TUs/SUs and the background data set should be equal. Discussion earlier in this review demonstrated that a minimum of 25 samples should be collected from each TU and for the reference background data set. If it could be proved that surface soil from each of the four on-site RBAs belong to the same statistical population, then the surface samples from each of the RBAs could be combined to form a single surface soil reference background data set. The same is true for the off-site RBA, if it could be proved that off-site RBA surface soils are from the same population as the on-site RBAs, they could be combined. However, there is insufficient evidence to support a single reference background population exists and in fact when plotted on a side-by-side box plot the analytical results for Radium-226 from surface soil samples collected from background reference areas during the historical investigation indicated very different variability and concentration levels per area. This was demonstrated in the Leidos March 2015 TM and is also depicted below.

Figure 1. Distribution of Ra-226 Off-Site Laboratory Measurements in picocuries per gram (pCi/g) in Background Reference Areas



A similar evaluation must be made prior to combining surface and subsurface sample results. Appendix A of the Work Plan acknowledges the expected disparity between the RBAs and states on page 3-1 *"Collections of fallout radionuclides are variable in soil at HPNS because of deposition and erosion, and from layering and mixing during placement of fill soil. Thus the concentrations of naturally occurring ubiquitous radionuclides may vary by location and depth."* It is unclear at this time, whether stratification of the soil in Parcel G also includes stratification of ROC concentrations within the RBAs, and if combining surface and subsurface sampling results within the individual RBAs or across the combined RBAs is appropriate for comparison to the soils in the TUS and SUs.

Discussion: Unless and until evidentiary data is collected to demonstrate that sample results from the four on-site RBAs can be combined to represent one reference background population, sufficient and an equal number of samples should be collected per RBA and per soil depth (surface and sub-surface) to allow for independent comparisons with on-site data. It is already noted in the Naval Draft Work Plan that the Bayview Area represents a different distribution of ROCs then the four on-site areas. Twenty-five surface soil samples are planned for collection at the Bayview Area; an equal number should be collected from the on-site RBA's.

SPECIFIC COMMENTS – PARCEL G

Executive Summary – Soil Investigations, 3rd paragraph, pg IV –

> *"A phased investigation approach is presented in this work plan that was designed to provide a high level of confidence that current site conditions either comply or do not comply with the Parcel G ROD RAO (Navy, 2009)."*
>
> As presented, the Draft Work Plan does not provide a means for quantifying the confidence associated with establishing compliance/non-compliance of Parcel G.
>
> > ***Discussion:*** Refer to **Parcel G – General Comments** of this TM for a detailed discussion. As stated earlier, it is widely recognized (MARSSIM 2002; NUREG-1505 Rev. 1) that statistical confidence is a critical component in establishing a defensible decision that a radiological site is in compliance. It is highly recommended that the Navy incorporate statistical testing into their data quality objectives (DQOs) so that conclusions drawn from this sampling effort can be substantiated and defended.

Introduction – fourth paragraph, pg. 1-1:

Additional information has been collected since the original Basewide Radiological Management Plan was designed. Historical data from background sampling can now provide a measure of expected variability for the soils on Parcel G and assist in the statistical design (sample sizes, statistical testing) of the Parcel G assessment and the reference background study. This historical information is relevant to the current conditions on the Parcel, including the Conceptual Site Model, and should be incorporated in the overall design of the assessment.

Section 3.4 - Radiological Investigation, third paragraph:

Please provide justification for how the number of TUs, to be excavated and undergo Phase 1 sampling, analysis, and scanning activities, was selected. What is the justification for only choosing 42 TUs of the 63 present in Parcel G? Similarly, how was the number of SUs determined for the Phase 1 investigation? The original assessment work conducted on Parcel G included the excavation of 100% of the TUs to meet the ROD RAO. Given the uncertainty of the historical sampling results and alleged data manipulation, what is the justification and supporting evidence for not excavating 100%?

Section 5.5 - Reference Background Area Soil Data:

It is unclear to this reviewer how the proposed comparison of background data and TUs will establish if an RBA is representative of on-site soils. As presented, this methodology is used for comparing duplicate sample results per the *Multi-Agency Radiological Analytical Protocols Manual* (MARLAP) not for comparison of two populations. Please provide a reference/guidance document that supports the use of this test to indicate a measure of representativeness when comparing two populations.

SPECIFIC COMMENTS - ESTABLISHING BACKGROUND - APPENDIX A

Section 2. Purpose and Data Quality Objectives, pg. 2-2, 3rd and 4th bullets –

The Kruskal-Wallis (KW) test is the non-parametric equivalent of a one-way analysis of variance (ANOVA) and is conducted on ranked data.

1) The KW test can be used on data sets which contain non-detect (ND) data if all the NDs are below the largest detected value.

   Recommendation: Add text that this assumption will be verified prior to running the KW test and provide an alternate statistical test or methodology if current/historical information or professional experience regarding the laboratory analysis of the ROCs indicate this assumption will not be met.

2) The null and alternate hypotheses of the KW test are based on the median of the data. In this case, the null hypothesis for each ROC is that the medians of all RBAs are equal, with the alternative hypothesis being that the median of at least one group is not equal to the medians of the other groups. It does not identify which, if any, of the RBAs are different from the others.

   Discussion: Please provide details on how the Navy will determine which RBA is statistically different from the others if the null hypothesis is rejected. Which post hoc test will be utilized?

Section 3.1.4 Number of Samples, 1st paragraph, pg 3-3 –

"The NRC Criteria for providing characterization of a complex site, found in United States Nuclear Regulatory Commission Regulation (NUREG) 1505 (NRC, 1998) is at least 100 samples from at least 5 distinct locations."

Discussion: Please provide a specific page number/citation from NUREG 1505 that supports the sample size and number of distinct locations cited in the above statement. This reviewer cannot

locate that recommendation within the NUREG cited. Further specification is required on how these numbers were derived: what assumptions were made, what inputs were used in the calculations, what data quality objective drives the sample size, cite a look-up table if one was used, etc.

Section 3.1.4 Number of Samples, Bulleted Sample Sizes, pg 3-3 –

The proposed sample sizes include 25 subsurface samples from each on-site RBA, 5 surface samples from each on-site RBA and 25 surface samples from the off-site Bayview RBA.

Discussion: Section 2 of the Draft Soil RBA Work Plan states that the RBAs, on-site and off-site, will be compared using a KW test. The fact that the Bayview Area sample size is five times greater than the on-site RBAs will compromise the outcome of the comparison. As stated earlier in this document, it is recommended that 25 surface soil samples be collected from each RBA unless statistical calculations can be provided that support 18 samples are sufficient to achieve power and a 99% confidence level for TU to RBA comparisons. This also provides a sufficient number of samples for between RBA comparisons. See the discussion for **Section 4.2.3** later in this document for further justification of increasing sample sizes within each on-site RBA.

Additionally, it is recommended that subsurface sampling also be conducted at the Bayview RBA to provide a comparison of the stratification of ROCs at depth with the on-site RBAs, visually and analytically.

Section 3.1.5.1 RBA-1 through RBA-4, 1st paragraph, pg 3-3 –

The geographical dimensions of the RBAs have been reduced significantly from the initially proposed dimensions in the February 2018 Draft Work Plan. Reducing the size of the area also reduces the probability of capturing the variability of ROC concentrations that is present across the originally selected RBA area.

Discussion: It appears that the areas were reduced to justify the collection of only 5 surface soil samples within each area so as not to saturate the new 20 foot by 20 foot square RBAs. As stated earlier in this TM, sample sizes per RBA per soil depth should be consistent with the sample sizes of the individual TUs/SUs. The original sizes of the RBAs should be re-established allowing for the greater number of samples to be collected at surface in a systematic way that represents the entire RBA not just a 20 x 20 foot area.

Section 4.2.3 Determine of Statistical Differences between Data Sets, 3rd paragraph, pg 4-4 –

"The RBA data sets will be compared to each other by applying the KW test, detailed in Section 13.2 of NUREG-1505 (NRC, 1998) and described in ***Section 4.1.3****, to determine whether the reference areas have similar or significantly different background levels."*

Discussion: Per NUREG-1505 Rev. 1 Table 13.5 shows that a minimum of 20 samples must be collected per RBA if 5 RBAs are selected for establishing background at an α=0.05 and power =97.6%. The Draft Work Plan presented by the Navy in February 2018 indicated that confidence levels would be set at α=0.01 and power=99%. To achieve those levels of confidence and power more than 30 samples would need to be collected per RBA per Table 13-5 of NUREG-1505, Rev. 1.

CONCLUSIONS

- Selected sample sizes for both the Soil Background Study and the assessment of Parcel G are inconsistent with MARSSIM and NUREG-1505 Rev. 1 guidance and need to be reconciled to assure defensible decisions are made regarding compliance/non-compliance for the soils located in Parcel G. Further documentation of sample size calculations that reflect the historical variability of the soils need to be included as part of the sample size justification. Although the original work conducted on Parcel G was based on a sample size of

n=18, which was defensible at the time given the lack of historical data, it must be re-evaluated to reflect the information that is now available to the Navy.

- Proposed comparisons of individual measurements of ROCs within TUs/SUs to mean/median background levels do not provide an associated level of statistical confidence, which is inconsistent with MARSSIM guidance. Either WRS/KW Tests or comparison to UTLs should be considered in addition to the proposed comparisons to provide a defensible decision of compliance/non-compliance.
- A clear statement of whether the Navy is following Scenario A or Scenario B of NUREG-1505 should be included in the Final Work Plan. In simpler terms:
    - Is the Navy assuming Parcel G still contains levels of ROCs that are not in compliance with the RAO and collecting data to prove otherwise (Scenario A)?
    - Or is the Navy assuming Parcel G soils are in compliance and collecting data to prove otherwise (Scenario B)?

  As stated earlier in this TM, there is a critical distinction between the two Scenarios. The burden of statistical proof is always on proving the original assumption is incorrect. As understood by this reviewer, the current Navy Draft Work Plan does not include a sufficient number of samples nor appropriate statistical testing to provide enough proof to refute the assumption that Parcel G soils are in compliance under Scenario B. The proposed Draft Work Plan is not a conservative plan that is protective of human or ecological health.

REFERENCES

Nuclear Regulatory Commission (NRC). *Nonparametric Statistical Methodology for the Design and Analysis of Final Status Decommissioning Surveys*. Interim Draft Report for Comment and Use. U.S. Nuclear Regulatory Commission Office of Nuclear Regulatory Research. NUREG-1505, Rev. 1. 1998.
https://www.nrc.gov/docs/ML0618/ML061870462.pdf (accessed on 2/18/2018)

MARLAP. *Multi-Agency Radiological Analytical Protocols Manual*. Volume I: Chapters 1-9 and Appendices A-E. NUREG-1576. EPA 402-B-04-001A. NTIS PB2004.105421. July 2004. (accessed on 7/27/2018)
https://www.epa.gov/radiation/marlap-manual-and-supporting-documents

MARLAP. *Multi-Agency Radiological Analytical Protocols Manual*. Volume II: Chapters 10-17 and Appendix F. NUREG-1576. EPA 402-B-04-001B. NTIS PB2004.105421. July 2004. (accessed on 7/27/2018)
https://www.epa.gov/radiation/marlap-manual-and-supporting-documents

MARLAP. *Multi-Agency Radiological Analytical Protocols Manual*. Volume III: Chapters 18-20 and Appendix G. NUREG-1576. EPA 402-B-04-001C. NTIS PB2004.105421. July 2004. (accessed on 7/27/2018)
https://www.epa.gov/radiation/marlap-manual-and-supporting-documents

MARSSIM. *Multi-Agency Radiation Survey and Site Investigation Manual (Revision 1)*. NUREG-1575 Rev. 1, EPA 402-R-97-016 Rev. 1, DOE EH-0624 Rev. 1, August 2002. (accessed March 12, 2018).
https://www.epa.gov/radiation/multi-agency-radiation-survey-and-site-investigation-manual-marssim

U.S. Environmental Protection Agency. ProUCL 5.1 User Guide. Statistical Software for Environmental Applications for Data Sets with and without Nondetect Observations. EPA/600/R-07/041. 2015.
https://www.epa.gov/land-research/proucl-version-5100-documentation-downloads

Visual Sample Plan (VSP). Pacific Northwest Laboratory. Version 7.1. 2018.
https://vsp.pnnl.gov

cc: Central File - WA # SERAS-106
Electronic File - I:/Archive/SERAS/106/D/TM/080218
Kevin Taylor, SERAS Program Manager (cover page only)