

RECEIVED

2003 JAN 28 A 9 01

CORRESPONDENCE CONTROL STATE OF COLORADO

Revised 12/02

CORRES CONTROL INCOMING LTR NO

00044RF03

DUE DATE ACTION

Bill Owens, Governor
Douglas H. Benevento, Acting Executive Director
Dedicated to protecting and improving the air and environment of the people of Colorado
4300 Cherry Creek Dr S
Denver, Colorado 80246-1530
Phone (303) 692-2000
TDD Line (303) 691-7700
Located in Centennial, Colorado
http://www.cdphe.state.co.us



Colorado Department of Public Health and Environment

January 17, 2003

Richard DiSalvo
Acting Assistant Administrator for Environment and Stewardship
U S Department of Energy-RFFO
10808 Highway 93, Unit A
Golden CO 80401-8200

RE: Draft Final Human Health Risk Assessment for the Solar Evaporation Ponds (December 2002)

Dear Mr DiSalvo

A Draft Final Human Health Risk Assessment for the Solar Evaporation Ponds was submitted as part of the December 2002 version of the Proposed Action Memorandum for IHSS #101. In response to our most recent previous comments on this risk assessment, this document appears to have undergone substantial revisions. Some of these changes have resulted in the text being less clear as to the exact methodology that was incorporated into the risk calculations.

Based on further discussions between the two regulatory agencies regarding distributional testing, the comments below describe two solutions. The first is to simply use the Shapiro-Wilk test for data sets less than 50 and the Shapiro-Francias test for sets >50 to determine normality/lognormality. The second resolution would be to input the data in the Pro-UCL software and use the software's recommendation for applicable distribution test.

If you have any questions regarding this approval, please contact Carl Spreng at 303-692-3358 or Jean McKenzie at 303-312-6258

Sincerely,

Handwritten signature of Steven H. Gunderson

Steven H Gunderson
RFA Project Coordinator
Colorado Department of Public Health and Environment



Table with columns: DIST, LTR, ENG. Rows include names like BOGNAR, E., CROCKETT, G., DECK, C. A., DEGENHART, K., DIETER, T. J., DIETERLE, S. E., FERRERA, D.W., FERRI, M.S., GERMAIN, A.L., GIACOMINI, J. J., ISOM, J. H., LINDSAY, D. C., LONG, J. W., LYLE, J. L., MARTINEZ, L.A., NAGEL, R. E., NORTH, K., PARKER, A.M., POWERS, K., RODGERS, A. D., SHELTON, D.C., SPEARS, M.S., TRICE, K.D., TUOR, N.R., WILLIAMS, J. L., BUTLER, L.

Reviewed for Addressee Corres Control RFP
1/28/03 Date By
Ref Ltr #
DOE ORDER # 5400-1

DOCUMENT CLASSIFICATION REVIEW WAYER PER CLASSIFICATION OFFICE

ADMIN RECORD 1101-A-000304

1/16

cc Scott Surovchak, DOE  
Dave Shelton, K-H  
Lane Butler, K-H  
Dan Miller, AGO

Steve Tarlton, RFOU  
Susan Chaki, CDPHE  
Admin Record, Bldg 850

**Colorado Department of Public Health & Environment**  
**Comments on**  
**Draft Final Human Health Risk Assessment of the Solar Evaporation Ponds**  
**(December 2002)**

---

**Section 2.1.2 – Power Calculations (Page 10)**

This section is quite confusing and should be rewritten for transparency as to the methods and findings

As presented in this report, it appears from Tables 2.2 a, b, & c, that datasets characterized by lognormal distributions require only one or two samples for each analyte. It is difficult to imagine any power test that would define one sample as being sufficient to adequately characterize a data set.

It is also unclear why there are two MARSSIM columns and two Lognormal columns presented in each table. This section requires additional detail as to what was done and what is being presented.

Additionally, this section appears to rely on results from distributional tests that have not yet been presented and assesses a subset of selected COCs that are discussed/identified later in the document. The rationale for the selection of these COCs in the power calculations is unclear (i.e., why some chemicals and not others?). The ones selected are not all final COCs.

Alternatively, the surface soil and liner data could be combined into one dataset for use in the risk assessment. By combining the data (no longer treating these two areas separately), there may be sufficient sampling data available to bypass these power calculations altogether and eliminate this entire section from the text.

**Table 2.2c (Page 13)**

This table has a footnote of “nc”. This footnote is not used anywhere in the table and should be removed.

**Section 2.3.1 – Essential Nutrients (Page 17)**

The word “Iris” should be capitalized to read “IRIS”. Additionally, this section is missing a conclusion statement that indicates which of the chemicals shown in Table 2.3 are eliminated as COCs based on this nutrient screen.

A couple of modifications are required in order to make Table 2.3 complete. Several of the abbreviations in Table 2.3 need to be defined in the footnotes (i.e., RDA/RDI/AI, UL2, ND). The RDA values shown in the table should be referenced as to their source. Lastly, since daily intakes (mg/day) are shown in the table (based on an assumed soil ingestion rate of 200 mg/day), the corresponding

s concentrations should be provided. The maximum is currently shown, so this would require adding a column for average site soil concentrations. Alternatively, if the maximum calculated intakes are below the RDA values, the comparison with RDA could be eliminated altogether.

#### Figure 2.10 (Page 20)

The figure should be updated to show all relevant PRG levels. For example, indicate that  $\text{risk} = 1\text{E}-06$  and  $\text{HQ} = 0.1$  are used for comparison with the maximum site concentration. Likewise, please adjust the hotspot evaluation to reflect use of a  $\text{risk} = 1\text{E}-05$  and  $\text{HQ} = 1.0$ . Additionally, please label the hotspot screen accordingly in the figure, so that readers realize the hotspot nature of the step.

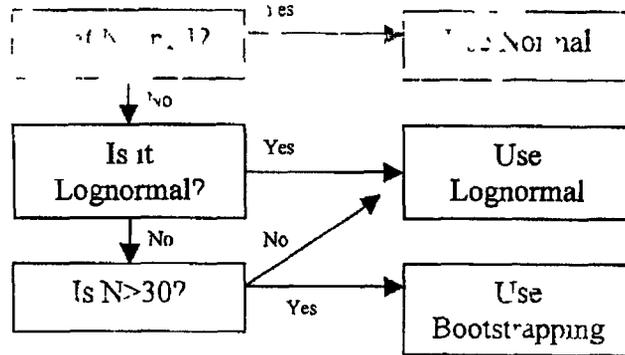
### Section 2.3.5 – Data Distribution Testing (Page 23)

The regulatory agencies share a concern over the selection of appropriate tests to determine data distributions. If the sample size is less than 50, the EPA QA/G-9 guidance recommends the use of the Shapiro-Wilk W test, wherever practicable. If the sample size is greater than 50, the guidance recommends using either the Filliben's statistic or the studentized range test. However, if critical values for these tests (for the specific sample size) are not available, then this guidance recommends implementing either Geary's test or the Lilliefors Kolmogorov-Smirnov test. Just because a software package contains multiple methods for determining a distribution, this does not imply that all methods are equally valid for a particular data set. I would suggest adhering to the guidance and using the most applicable test for each data set being evaluated.

As currently written, there are many potential problems with the approach provided in the December 2002 SEP reports. The EPA QA/G-9 report identifies limitations to several of these methods that are included. For example, Geary's test is recommended for data sets with greater than 50 samples. The SEP report uses this test for sample sizes of 15 in the liner material. Additionally, G-9 indicates that "this test does not perform as well as the Shapiro-Wilk test or the studentized range test." However, it appears that the SEP report has given equal weight to this test, despite these limitations.

Likewise, the Coefficient of Variation test (CV) can be used to determine that a curve is not normal (i.e.,  $\text{CV} > 1$ ). However, as clearly stated in G-9, this method should not be used to conclude that data can be modeled with a normal curve if the CV is less than 1. This method should be applied only to quickly discard an assumption of normality, and not to conclude normality. The SEP report attempts to use this method contrary to its intended use, by assuming normality based on test statistics. The only result that is applicable in the report tables under the heading "CV" is the "No" value (discarding normality). The "Yes" values should be removed and not considered in the evaluation process.

The agencies also share a concern regarding the decision to treat the data as non-parametric when the results indicate that both the normal and lognormal distribution apply. Preferably, if the data pass the test for normality, there is no reason to test for log-normality as shown in the following flowchart of decision logic.



There are two solutions to proceeding with the distributional testing. The first is to simply use the Shapiro-Wilk test for data sets less than 50 and the Shapiro-Francia's test for sets >50 to determine normality/lognormality according to the flowchart provided above. For datasets containing negative values, first test for normality and if the analyte fails this test, move directly to bootstrapping methodology. The second resolution would be to input the data in the Pro-UCL software and use the software's recommendation for applicable distribution test.

**Table 2.9 (Page 25)**

The sample counts shown in the left hand column need to be updated to reflect the number of samples collected in background surface soil samples, rather than the number of samples collected in the liner, as is currently shown.

**Surface Soils Data Evaluation (Page 25)**

The text indicates that surface soil data were evaluated for each PCOC with a maximum above the WRW PRG. Results from distributional testing are shown for 10 analytes in surface soil, however, 12 chemicals were found to fail the PRG screen. Please include distributional testing for Benzo(a)pyrene and Dibenzo(a,h)anthracene.

**Subsurface Soils Data Evaluation (Page 26)**

The text indicates that distributions were evaluated for all PCOCs retained in the PRG screen. This section leaves out the result for distributional testing of benzo(a)pyrene, a chemical which was retained in the PRG screen.

**Table 2.13 (Page 27)**

The abbreviation "na" is given for Am-241 under the S/K test for normality. Please footnote the table with a relevant description.

**Surface Soils (Page 31)**

The first full sentence on this page refers the reader to a non-existent Section 2.8 for further discussion.

**Application of Professional Judgment (Page 31)**

For benzo(a)pyrene the 95UCL, calculated using a bootstrap methodology, is compared to the WRW PRG value. The document does not show results from the distributional testing for this analyte that supports use of the nonparametric bootstrapping statistics. The underlying distribution (normal, lognormal, nonparametric) for this data set should be shown prior to calculating a 95UCL value.

**Tables 2.15 to 2.17 (Page 33-35)**

The purpose of including these tables is unclear. The paragraph immediately preceding Table 2.15 states:

"Only compounds listed in ALF were assessed for the risk assessment, per agreement. All analytes listed in ALF had toxicity factors. Tables 2.15 through 2.17 list analyte[s] with no PRGs in ALF."

It is assumed that this is a listing of non-ALF chemicals that will not be addressed by this risk assessment or in this PAM. There is, therefore, no reason to present these tables in the body of the risk assessment, as it may lead to confusion. They may be better placed as an appendix, if deemed necessary. It is assumed that these individual records are already presented in one of the appendices that lists comprehensive summary statistics for all analytes by media.

Additionally, several of these analytes (calcium, magnesium, potassium, silicon, and sodium) have already been addressed in Section 2.3.1. The discussion of whether or not these are ALF-chemicals is relevant to their inclusion.

However, if toxicity factors exist for some of these analytes found in the soil at the solar ponds, then PRGs/ALs could be calculated. For example, thallium and ethyl acetate are listed in these tables, but have oral RfDs in IRIS. Additionally, HEAST has numerous values that could be used to derive PRGs for the radionuclides.

**Tables 2.18 to 2.20 (Pages 36 & 37)**

There is a footnote on these tables that indicates the data were calculated using bootstrap resampling methodology. It is unclear as to what in these tables would have been derived using bootstrapping, since they present only summaries of count and detection frequency for each COC.

6

**Section 3.1 – Future On-site Land Use (Page 38)**

This section contains a statement about the presence of Preble's habitat at the site. If this is included, then the risk assessment should explain how this will be addressed, rather than ignoring this land use when identifying exposure pathways and receptors.

**Future On-site Land Use (Page 38)**

The text indicates that the worker will spend 90% of each day outdoors across the Site with an emphasis near the watershed areas. This language refers to a concept that may be part of upcoming CRA discussion, but that has not yet been agreed to. For example, someone conducting a prairie dog survey would not be expected to preferentially visit the watershed areas. For the purposes of the SEP risk assessment, it is assumed that random exposure may occur across the entire AOC, with no preference towards a specific sub-location. The sentence should be removed.

**Section 3.3 – Exposure Scenarios (Page 42)**

The receptors for evaluation in the CRA are still undergoing negotiation. Therefore, it is premature to state that risks to off-site receptors "will be addressed in the Site CRA."

**Section 3.4 – Exposure Point Concentrations (Page 45)**

In the first paragraph, the text states that problems arise with assuming lognormality when data are not lognormally distributed. Please expand this discussion by referencing the types of problems that are known to occur, rather than just making this simple statement.

**Section 3.4 – Exposure Point Concentrations (Page 46)**

In the first paragraph on page 46, please provide more details on the bootstrapping methods that were used for the SEP risk assessment. There are many different variations of the bootstrap method that have been developed. According to several sources, the number of bootstrap samples appropriate for developing reliable confidence limits depends on the statistic of interest and the acceptable error in the interval. A minimum of 1,000 replicates is generally recommended and was used for the SEP risk assessment. Confirm that 1,000 replicates was sufficient to characterize the statistic of interest.

**Table 3.3 (Page 47)**

It would be much more helpful to have a table that shows the underlying data distribution determined for each chemical, as well as a column summarizing the final EPC value used for the risk calculations. In some cases (e.g., Americium), the EPC value defaults to the maximum detected concentration. This is not always clear to the reader and a summary column would eliminate potential confusion.

Uranium is listed in this table, but was not brought through the COC selection process. Based on the final risk values shown at the end of the document, uranium

could very well end up being eliminated *a priori* as a COC. This analyte should follow the same procedures established for the other analytes in this risk assessment.

**Table 5.1 & 5.2 (Pages 56 & 57)**

Why don't the HI values presented in Table 5.1 (by medium) match the HI values shown in Table 5.2? For example, Table 5.1 shows a dermal HI for surface soil of 0.008, whereas Table 5.2 has a summed dermal HI for surface soil of 0.006.

**Section 5.4.2 (Page 61, First Paragraph)**

Change "assign" to "assigned"

The text states that "distributional testing was also conducted for individual surface soil COCs using the Shapiro-Wilk test on the data and Ln-transformed data." This statement does not accurately reflect what was presented in earlier portions of the text.

**Section 5.4.2 (Page 61, Third Paragraph)**

This paragraph indicates that, "The Bootstrap method was used to calculate UCLs for the SEP Risk Assessment." This statement is confusing, because in fact, the bootstrap method was only used to calculate UCLs for those analytes that were determined to have non-parametric distributions and sampling sizes greater than 30. If this section is designed to compare UCLs calculated via multiple methods, simply state that. Do not confuse the reader by implying that the presented values were used to calculate site risks.

The table is footnoted that the lognormal statistics were not applicable for some of the analytes, since their distributions were not lognormal at the 0.05 level. It is unclear which of the five tests this determination was based on. This same evaluation was not performed for the assumption of normality. Of the chemicals presented, none were determined to have normal distributions in Table 2.10, therefore, comparing the bootstrap or geostatistics value to this "normal" UCL95 value and declaring a similarity is inappropriate, since it has already been determined that the assumption of normality for this dataset is inappropriate. In fact, since the default is to assume a lognormal assumption, it seems more appropriate to compare the bootstrap and geostatistical values to the lognormal UCL.

**EPA Comments  
on the 2<sup>nd</sup> Draft Final Human Health Risk Assessment  
dated December 2002 for the Solar Evaporation Ponds**

EPA Toxicologic Review

**General Comment**

1

Oddly enough, this revision appears to be further away from completion than the previous draft. There are still a number of errors in the report which need to be corrected. However, my major concern is with the process used to determine distribution shapes and exposure point concentration terms for the contaminants of concern (COCs). Instead of following the simpler, straight forward approach recommended by EPA and CDPHE in previous meetings and memorandums, the document uses an overly complex and in my opinion, unnecessary approach for testing distribution shapes, and presents the results in an incomplete and confusing manner. In addition, the credibility of the document is not enhanced by numerous sections criticizing the methodologies recommended by EPA and CDPHE. My specific comments are as follows:

**Specific Comments**

**1. Page 24, Section 2.3.5 Data Distribution Testing**

Instead of using one statistical test for normality, the authors used five. If DOE chooses to spend the time and resources pursuing this level of detail, I have no problem. However, each test was intended to be used for different types of data sets with different detection limits, distributions, sample size, etc. This is not explained in the document. As written, the document gives equal weight to all of the tests. The second bullet on page 24 states that two or more "no" results for the tests indicates that the data did not conform to the distribution being tested. This is not a very useful or technically accurate guideline if the two tests with the "no" results are inappropriate for the specific data sets being tested. This section needs to be revised. I've outlined two approaches below which would result in a document satisfactory to both EPA and CDPHE.

- Simplify the approach for testing distribution shapes as recommended by EPA and CDPHE in previous meetings. Specifically, use one test for normality for samples sizes less than 50, such as the Shapiro-Wilks, and use one test for samples sizes greater than 50, such as the Shapiro-Francia test. Follow the diagram provided by CDPHE in interpreting the results of the test. Specifically, if the results indicate the data set is normal, then use the normal distribution. If the results are not normal, then test for lognormality. If the results indicate lognormality, then assume lognormality. If the results are neither normal, nor lognormal, then assume they are non-parametric. Sample sizes less than 30 are to be treated as lognormally distributed.

- ☛ Use a more complex approach utilizing multiple tests for testing distribution shapes. The major problem with the current draft is that multiple tests are used, however, no effort was made to evaluate the appropriateness of the tests to the site-specific data sets. We recommend that DOE use the new EPA ProUCL (Version 2.1) software. The software will run multiple tests evaluating distribution shapes, recommend the appropriate distribution for each specific data set and calculate the exposure point concentration based on the recommended distribution.

The third bullet on page 24 states that the data are to be treated as non-parametric when the results indicate that both the normal and lognormal distribution apply. We don't agree with this decision rule. If the data are shown to be normal, then assume a normal distribution. Do not proceed with any further testing. Data sets for which either the normal or lognormal distributions fit are exhibiting low variability. The choice of the normal distribution is not only reasonable, but it is also to the advantage of the regulated party.

The 4<sup>th</sup> bullet on page 24 states that radiological data with zero and negative concentrations are considered normal or non-parametric. I agree with this approach.

## **2 Page 25, Surface Soil Data Evaluation**

The 2<sup>nd</sup> paragraph states that none of the surface soil COCS were classified as normal or lognormal. Yet, Table 2.10 shows arsenic as being normally distributed. This sentence needs to be revised.

## **3 Page 44, 4<sup>th</sup> bullet**

The fourth bullet references a 1991 EPA guidance which is outdated and has been superseded. The paragraph should be revised to explain that the equations and parameters for the radionuclide external exposure pathway are taken from the October 2000 *Soil Screening Guidance for Radionuclides User's Guide*.

Also, note that the mass loading factor used in Tables 3.1 and 3.2 has changed back to the 50<sup>th</sup> percentile value. In the last draft final we reviewed, the mass loading value was consistent with the RSAL Task 3 report. Although I think this is a more realistic value, you and Carl should be aware that it will be different and be prepared to respond to the public.

## **4 Page 45, Section 3.4 Exposure Point Concentrations**

The first sentence under this section appears to have escaped the revisionists' pen from earlier rounds. The sentence should state that the exposure point concentration is the 95 UCL on the arithmetic mean. Period. The clause "assuming normality" is an error and should be deleted.

The 7<sup>th</sup> line down in the first paragraph is written awkwardly, if not erroneously. It states that a geometric mean and 95UCL are calculated when data distribution is demonstrated to be lognormal. This sentence should be revised to state that the geometric mean and 95% UCL of the transformed data are calculated when it is demonstrated that the calculated values of exposure point concentration are lognormal. Lognormal data sets should be re-checked because the geometric mean and 95% UCL of the other results are marked differences.

#### **5 Page 46, 1<sup>st</sup> full paragraph**

*Note: EPA and CDPHE are requesting that the liner data be combined with the surface soil data within the SEP. In that event, this comment is no longer specific to the liner data, but this should not preclude a revision to Table 3.3 to clarify the text.*

The 9<sup>th</sup> line down in the 1<sup>st</sup> full paragraph states that lognormality was assumed for all final COCs in liner material based on direction from EPA, CDPHE to assume lognormality for all data sets with less than 30 samples. This is correct. However, this section is written in a confusing manner and it is not easy to see what was actually used as a concentration point exposure term. The easiest way to resolve the confusion is to revise Table 3.3 in Section 3.4 adding columns to clearly show the distribution shape assumption used in the calculation for each analyte as well as the actual exposure point concentration term chosen. A footnote should be added to the table explaining the policy decision.

The last paragraph on page 46 contains a typo in the 5<sup>th</sup> line. "UCLf or" should be "UCL for".

#### **6 Page 47**

The 1<sup>st</sup> paragraph explains that the maximum value of 8.1 was used as the EPC for americium instead of the 95% UCL. This should be shown on Table 3.3 in an additional column.

The last sentence on page 47 states that all COCs in surface soil had non-parametric distributions. Table 2-10 on page 25 shows differently. The sentence should be revised.

#### **7. Page 49, Table 3.4**

The risk equation for external radiation risk is missing the gamma shielding factor.

**8. Page 53, Table 4.1, Toxicity Factors**

The 3<sup>rd</sup> column in Table 4.1, labeled "DAF fraction" appears to confuse a dermal absorption fraction, which is a variable in the exposure equation, with a gastrointestinal absorption efficiencies which are used to adjust the oral toxicity values. Exhibit 4-1 in EPA's 2002 Dermal Guidance provides absorption efficiencies of 2.5% for both cadmium and chromium. These values should be used to adjust the toxicity values. If an absorption efficiency is 50% or greater, the toxicity value should not be adjusted.

**9. Page 61, 2<sup>nd</sup> paragraph**

The second paragraph implies that the methods recommended by EPA and CDPHE to test for distribution and calculate the exposure point concentration term are inappropriate. This does not add to the credibility of the report and should be deleted from text.

**EPA Statistical Review:**

1. **Section 1.0, Page 1, last sentence.** Ecological risk is not addressed in this human health risk assessment (HHRA), but will be addressed in the future. The proposed remedial action is to bulldoze the earthen berms and cover the asphalt liners of the Solar Evaporation Ponds (SEP) as soon as regulatory is obtained. Regulatory approval would follow a finding by the Colorado Department of Public Health and the Environment (CDPHE) and U.S. Environmental Protection Agency (EPA) that human health risks are not significant and no excavation of contaminated soil or liner material is required. The intended remedial action does not address the possible outcomes of the ecological risk assessment, including the possibility of significant risk to an ecological receptor. These issues should be addressed.
2. **Section 1.1, Page 2.** This section states, "Contaminated liquids apparently infiltrated into subsurface soil." This statement is unclear. The word "apparently" should be deleted.
3. **Section 2.1.1, Page 8.** Only 15 asphalt liner samples were collected for metals and radionuclides. Yet in the summary on page 10, data quantity is stated to be acceptable for HHRA purposes. It is not clear how 15 samples can meet the data quality objective (DQO). Liner data should be combined with surface soil data.
4. **Section 2.1.2, Page 10.** A previous submittal dated September 11, 2002, included responses to CDPHE and EPA Comments on the Solar Evaporation Ponds Project. In that document, 66 samples per the Gilbert equation power calculation (13.23) was determined as the minimum sample size necessary to characterize surface soil radionuclides. This Gilbert equation uses the median rather than the 95 upper confidence level (UCL) and therefore is not conservative.

The new method, that calculates relative errors as the difference between the PRG or action level and the mean or 95UCL, results in only one sample being required. The MARSSIM method results in 13 samples being required. An order of magnitude difference between these two results is cause for skepticism regarding this approach.

It appears that the rationale given in the third HRA is an attempt to justify the collection of only 15 samples. The present strategy attempts to discount the need for calculating the 95UCL to any precision by demonstrating the large difference between its value and the preliminary remediation goal (PRG) associated with a risk of one in hundred thousand (1E-05). However, the strategy fails, because not enough samples have been analyzed to reliably estimate the 95UCL.

As an example, if the Gilbert equation requires 66 samples to estimate the median of a lognormal distribution with a relative error of 10 percent and confidence of 95 percent, the lognormal variance can be back calculated as 0.156. Substituting this variance and assuming the number of samples required is 1, the Gilbert equation can be solved for  $d$ , the relative percent error (same as  $\beta * 100$  percent) allowed. The answer is 116.9 or ~117%. Because power =  $1 - \beta = 1 - 0.17 = 0.83$ , one sample has no power to determine the 95UCL.

All power calculations should be recalculated in a manner consistent with standard statistical practice per the previous submittal but on individual PCOCs.

Also, an alternative approach should be considered. If the Liner data are combined with surface soil data typically resulting in ~85 samples per PCOC, the DQO will be considered achieved and Section 2.1.2 Power Calculations should be removed from the risk assessment.

- 5 **Figure 2.6, Page 20.** It is unclear why potential contaminants of concern (PCOCs) at concentrations corresponding to less than 3E-05 were deleted from further analysis if the target risk level is 1E-05. The rationale for eliminating these PCOCs from the analysis should be provided.
- 6 **Section 2.3.5, Pages 23-27.** Using five tests of normality in a weight-of-evidence approach is acceptable. However, requiring 4 of 5 tests to be "yes" is too stringent, three of five tests should be acceptable. Some of the normality tests chosen are questionable in terms of power and sample size. As such, the tests should not be given equal weight as they have been.

The third bullet on page 24 is incorrect as stated, and the final distribution in Tables 2.8, 2.9 and 2.11 are also incorrect in many instances. These items should be corrected.

An alternative simpler approach is recommended. One test, the Shapiro-Wilk W test is sufficiently stringent to be acceptable by itself. Many standard statistical packages give the probability ( $p$ ) of fit of the Shapiro-Wilk W test. Also, the Shapiro-Francia test is an extension of the Shapiro-Wilk test good for sample sizes up to 2000. The Lilliefors test is also acceptable for sample sizes  $> 50$ . ProUCL v 2.1, freeware, can perform the Shapiro-Wilk W and Lilliefors tests.

- 7 **Section 2.3.7, Pages 33 to 35 and Tables 2.15 to 2.17.** It is stated that only compounds listed in ALF were assessed for the HHRA. The text does not include which compounds were dropped or why, if all parameters are available to calculate risk, the compounds were dropped. This information should be included.
- 8 **Section 3.4, Page 46, First Paragraph.** Most bootstrap texts mention 1,000 or 2,000 replications as sufficient. Five thousand, 10,000 or more would be better, but results would probably be only slightly greater. An adequate number of replications may be found with a decision rule such that doubling the number of replications results in a numerical change of say only 1 percent or whatever is deemed acceptable. There is no hard and fast rule on the percentage chosen.

A check that 1,000 replications are sufficient should be conducted by performing more until the final result agrees to within 1% of the previous result.

- 9 **Section 3.4, Section 3.4, Page 47, first sentence and Table 3.3.** The text states 8.1 pCi/g for Americium-241 was used to calculate risk. However, the maximum value in Table 3.3 is given as 8.19 pCi/g which rounds to 8.2 pCi/g. Only the final answer in all risk equations should be rounded in accordance with standard practice.
- 10 **Section 5.4, Pages 58-59, Section 5.4.3, Page 65 and Section 6.0, Page 66, Last Bullet.** This section and the summary stress the uncertainties associated with conservative assumptions in the HHRA and states that actual risks may be lower. However, chemicals that are not on the Action Levels and Standards Framework for Surface Water, Ground Water, and Soil (ALF) list are not evaluated. In addition, chemicals without toxicity values are not evaluated. The HHRA states that this adds a degree of uncertainty to the risk assessment, failing to acknowledge that this may result in actual risks being higher.

A statement should be added that acknowledges risks may be higher as a result of not evaluating chemicals.

- 11 **Section 5.4, Page 61, second paragraph and Table 5.5.** This paragraph states that "good spatial correlation" occurred among the COC data that, in turn, dictates use of geostatistics. This conclusion negates everything presented previously in this HHRA where classical statistics were employed. Considering that the pattern of historical contamination was probably randomly located spills as opposed to continuously distributed spills, both classical statistics and spatial statistics have some merit, but neither is perfect.

The three paragraphs beginning with "In addition, a Geostatistical Spatial Analysis" leading to the subsection Mass Loading and Air Exposure Concentrations should be removed from the risk assessment. A discussion of the differences in UCLs obtained by the different statistical methods could be included in the Uncertainties section of the risk assessment. In any case, since none of the COCs in Table 5.5 were normally distributed as shown in Table 2.10, the normal column in the table is irrelevant and should be removed.

The three paragraphs beginning with "In addition, a Geostatistical Spatial Analysis " leading to the subsection Mass Loading and Air Exposure Concentrations should be removed from the risk assessment. A discussion of the differences in UCLs obtained by the different statistical methods could be included in the Uncertainties section of the risk assessment, in any case, since none of the COCs in Table 5.5 were normally distributed as shown in Table 2.10, the normal column in the table is irrelevant and should be removed.

### **Additional Comments on Appendix A**

Time constraints necessitated that review of Appendix A be confined to one important COC, Americium-241. There are unresolved issues with the database being used for the risk assessment as enumerated below. Numerous apparent errors were found suggesting a disconnect between Appendix A and the risk assessment text and tables.

#### **1 Appendix A, Table A.3 - Solar Evaporation Ponds AOC Analytical Results for Surface Soils Radionuclides, Page 1 and Table A.13a**

There are 69 Americium-241 surface soil samples in the Appendix A database. Gannett Fleming (GF) also prepared a database for the Rocky Flats SEP. The GF database has 54 Americium-241 surface soil samples. The largest Americium-241 concentration reported in Appendix A is 130 pCi/g. The largest Americium-241 concentration reported in the GF database is 220 pCi/g.

These data discrepancies should be explained before any confidence can be placed in the Appendix A database.

Assuming the database in Appendix A is correct, Americium-241 is lognormally distributed according to the Shapiro-Francia test ( $p = 0.07$ ). The arithmetic mean is 8.69 pCi/g as reported in Table A.13a. But Table 3.3 in the RA reports 9.11 pCi/g as the mean.

All numerical inconsistencies between Appendix A and the risk assessment text should be resolved. A column should be added to Table 3.3 citing the source of the 95UCL concentration (normal, lognormal, bootstrap).

#### **2 Appendix A, Table A.5 Solar Evaporation Ponds AOC - Analytical Results for Liner - Radionuclides, page 1 and Appendix A, Table A.13b Solar Evaporation Ponds AOC - Summary Statistics for Detected Analytes in Liner**

Six of 15 Americium-241 Liner samples show a lab result qualifier as "<" which would seem to indicate the samples are non-detects. The magnitude of these results also suggests these samples are non-detects with two levels of censoring (0.003 and 0.005 pCi/g). However, Table A.13b indicates 100% detections for these 15 Americium-241 Liner samples. Also, Table 2.19 page 37 of the RA indicates these data are non-detects by reporting a detection frequency of 60%. These discrepancies should be corrected.

The same discrepancies occur for Liner COCs other than Americium-241. These should also be corrected.

3 Appendix A, Table A.24-A 79

No Liner results were found in these background comparison tables. The use of surface soil background as "surrogate" Liner background as stated on page 28 of the risk assessment is inappropriate and unacceptable. However, since the Liner samples are to be combined with the surface soil samples to meet the DQO, the use of background surface soil for the background comparison will then be acceptable.

4 Appendix A

No background data were found. Reference is made to documents containing background data in the risk assessment. All data used in the risk assessment should appear somewhere in the risk assessment or appendices.

A Table of Contents should be added to Appendix A to help the reader more easily access the data.