4-21-83 E

DOCUMENT DOES NOT CONTAIN ECI
Reviewed by Date 4/29/92

Battelle
Pacific Northwest Laboratories
Battelle Boulevard
Richland, Washington 99352
Telephone (509) 946-2104
Telex 32-6345

409806

September 22, 1976

REPOSITORY P.N.N.L.
COLLECTION Marshall Islands
BOX No. 5685
FOLDER Enewetak 1976

Mr. Tom McCraw
U.S. Energy Research and
Development Administration
Washington, D.C. 20545

Dear Tom:

This letter is in response to your request for guidance on the number of samples required for the proposed clean-up survey on the Enewetak atoll. I begin with some general comments then discuss specifically the questions you distributed at the meeting in Joe Deal's office on July 29, 1976. There is also an appendix to illustrate the computation of certain confidence limits using $^{239-240}$Pu soil data from the island of Janet. This letter has benefited from comments and suggestions by other statisticians at BNW (Drs. Lee Eberhardt, Tony Olsen, and Pam Doctor).

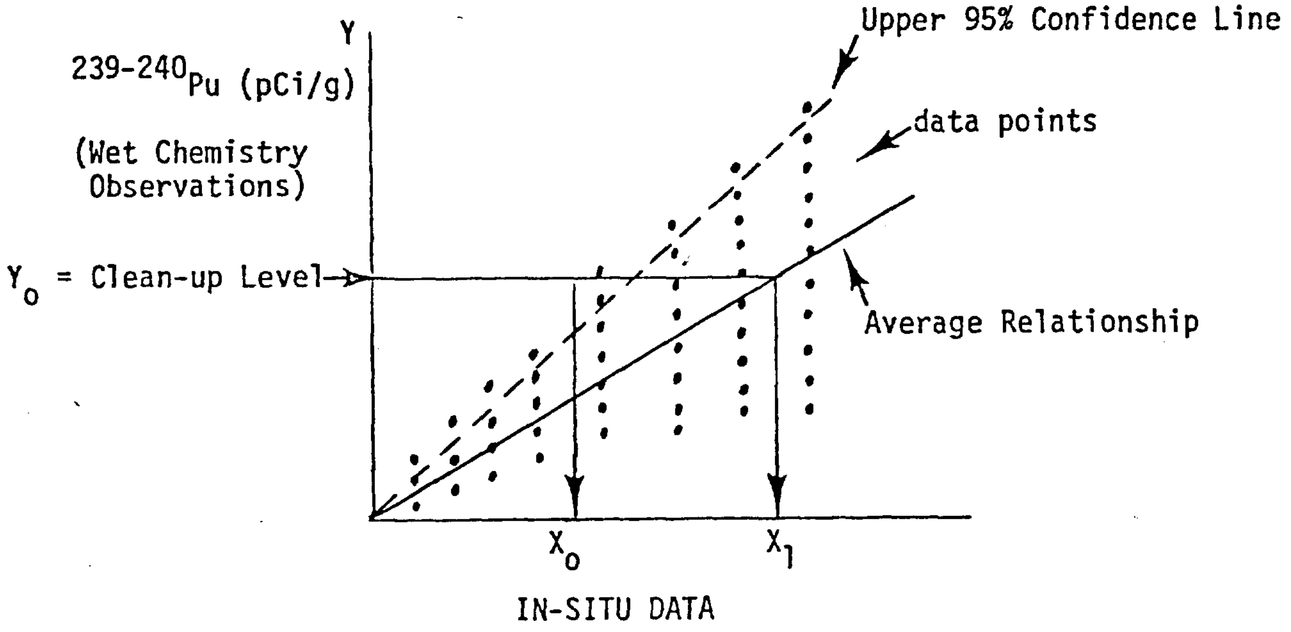
The number of samples will depend in part on how well the portable Ge(Li) counter performs in the field, i.e. on how accurately the Ge(Li) readings relate to the amount of plutonium in soil. It will also depend on the statistical design used in the field and on whether it is decided that a contour map of plutonium concentrations is a major goal or whether probability statements about mean concentrations are preferred. Contouring calls for a systematic (uniformly spaced) sampling scheme, while probability statements require random sampling within sub-areas of an island. Probably it would be desirable to use some kind of sequential sampling scheme, in which results of an initial set of samples are used to decide whether a given area should be (a) considered "clean" (below some standard level), (b) cleaned up, or (c) whether additional samples should be taken before a decision is made. Such a scheme is likely to require continued attention by someone with statistical training, but may be expected to reduce the amount of sampling required.

If contouring is used, Dr. Delfiner of the Centre de Morphologie Mathematique, Fontainebleau, France should be consulted on this question of the number of samples required. Dr. Delfiner is knowledgable on "kriging" (a contouring method), and he may be helping Bruce Church set up the technique for use on the islands. We understand that arrangements are being made for Dr. Delfiner to be in Las Vegas for 3 weeks in October and again in November to install his kriging routine on REECo's computer.

The question of whether In-Situ measurements, soil samples, or both should be used for deciding whether an area or island should be cleaned-up requires further discussion. To answer this question we need to know whether In-Situ or soil sample data are more indicative of future health risk of inhabitants. Is long term health a function of an average (integrated) measure of exposure such as obtained by an In-Situ device, or is it more a function of peak plutonium concentrations from soil aliquots? Also, the use of In-Situ measurements in the cleanup determination implies we need to determine the relationship of these readings to soil concentrations. This can only be done under field conditions. I would suggest a number of In-Situ measurements be made at different locations. At each location the total soil scanned by the device should be carefully collected and mixed and a number of aliquots be analyzed for plutonium. In this way a calibration equation relating In-Situ and soil sample concentrations can be estimated. This will no doubt need to be repeated for different islands or portions of islands since the calibration relationship may not be the same for all areas. If the decision to cleanup is made primarily on the basis of In-Situ measurements then the calibration information is necessary in order to relate to the cleanup criteria which, presumably, will be stated in terms of plutonium concentrations in soil samples.

This reflects a basic decision needed before a sampling plan is selected. If clean-up decisions are to be based on wet-chemistry determinationson soil samples then the In-Situ device may serve only as a means of reducing the number of analyses needed. In any case, we suppose some chemical determination will be required for calibration of the device.

Let us assume that the In-Situ measurements are related linearly to the average Pu concentration in the surface soil scanned by the In-Situ device. For example, if the Pu/Am ratio is constant then the data should look something like a linear "average relationship" through the origin as indicated in the plot below.



The variability in Pu concentrations would probably be greater for high In-Situ readings than for low readings (as indicated in the diagram). If Y_0 is the level of Pu in soil signifying clean-up, this corresponds to an average In-Situ reading X_1. But the data in the diagram indicates individual Pu readings considerably greater than Y_0 for In-Situ reading X_1. Hence, if the In-Situ device is used to meet the clean-up criteria in terms of Pu concentrations, the level of In-Situ indicating clean-up should be less than X_1. One candidate is the value of the In-Situ measurement (X_0 in the diagram) such that the upper (one-sided) 95% confidence on average Pu concentration is Y_0. An alternative approach would be that level of In-Situ reading such that some large percent (P) of the Pu concentrations associated with that In-Situ level are less than Y_0 with probability $1-\alpha$. The main point here is that if clean-up is to be based on In-Situ measurements, the level of In-Situ measurement indicating clean-up should probably be lower than indicated by the <u>average</u> linear relationship.

In the remainder of this letter I have addressed the five questions you handed out at the meeting with Roger Ray, Paul Dunaway, and others in Joe Deal's office on July 29, 1976. Hopefully, this discussion will help clarify some of the different kinds of statistical probability statements that can be made based on sample results. I direct your attention particularly to the discussion of "acceptance sampling" for Question 3. This seems to be a much more satisfactory approach than using average Pu concentrations for deciding whether an island needs to be cleaned up. There are a good many details that would need to be worked out for actual field application in connection with kriging, but these need to be explored with someone like Dr. Delfiner. A table of sample sizes required to meet various probability criteria is included in the section dealing with Question 3 for the simplest (nonsequential) sampling design. The number of samples would probably be less for a sequential design.

<u>Question 1</u>: Over what area or areas should Pu-in-soil measurements be averaged:

a. In-Situ measurements?
b. Soil sampling?

The answer to this question depends in part on the variability present from sample to sample, the spacing of samples, whether any trends are present and perhaps most importantly on how the health standards (cleanup criteria) are formulated. If there are no trends and the variability between samples is relatively small, then the area over which samples are averaged can be large. However, if strong trends are present (such as near GZ for example), it would be important to define these fairly precisely. In that case rather few if any areas might be averaged. Presumably In-Situ measurements would need less (if any) averaging than plutonium concentrations in soil samples since each such In-Situ measurement is itself an average of the Americium activity in the area scanned by the detector.

It is clear that averaging plutonium concentrations will tend to reduce the apparent health risk since the peak concentrations get averaged in with the lower concentrations. This is not, however, a justification for averaging. What we need to know is what average or metric is the best indicator of future health risk to persons inhabiting the area. Guidance from resuspension and radionuclide cycling studies is needed here.

Question 2: To what areas should the Pu cleanup criteria, 40 pCi/g and 400 pCi/g, be applied?

This seems to be a restatement of Question 1. Again, the answer depends on how concentrations for the various size areas are related to health. If this were known and we had some idea of trends and variability over space, we would be in a better position to answer this question.

Question 3: Looking at past survey results compared with the cleanup criteria, which islands need cleanup? What levels of assurance that the criteria are met without cleanup are reasonable and attainable?

A. There are a number of probability statements that can be made based on survey data. These include (1) a one-sided upper confidence limit on the true (unknown) average Pu concentration, and (2) a one-sided upper confidence limit on a percentile of the population. For this latter case, using the 95th percentile for $\alpha = .01$ as an example, we could construct, e.g., an upper $100(1-\alpha) = 99\%$ confidence limit on the concentration level below which 95% of the soil concentrations on the island lie. A third type of interval that appears particularly useful is a one-sided upper confidence limit on the proportion of soil concentrations that fall below the cleanup specification level (this level is denoted here by L). These three kinds of limits are illustrated in an attached supplement to this letter using the $^{239-240}$Pu data collected on Janet during the 1972 Enewetak survey. We might say at this point, however, that confidence limits on average values (number 1 above) are usually computed on the assumption the data are themselves normally distributed or that the estimated mean is normally distributed. Since Pu concentrations tend to have skewed distributions similar to the lognormal, the usual procedures are sometimes modified by first transforming the data to logs, computing the limits in log scale, then transforming the limits back to the original scale. Alternatively, nonparametric or "distribution-free" limits can be computed. These latter limits are valid no matter what the underlying statistical distribution, but the one-sided limits will be higher (or wider for 2-sided limits) than if a specific distribution such as the normal or lognormal is assumed. We note, however, that limits on percentiles and proportions (items 2 and 3 above) do not require any assumptions about the underlying statistical distribution. The several approaches mentioned above are illustrated in the Supplement.

B. The question of whether to cleanup an island or part of an island can be put in a hypothesis testing framework. In particular, what is known as "acceptance sampling" appears to be a useful approach since there is no need to make any assumptions (normal, lognormal, etc.) about the statistical distribution of the data. The basic idea is to specify (1) an activity level, say L, above which cleanup is indicated, (2) a proportion (p_1) of samples with activities greater than L that is acceptable, (3) a proportion (p_2) of samples with activities greater than L that is not acceptable, (4) the allowable risk (α) of concluding that cleanup is necessary when it really isn't, and (5) the risk (β) of concluding that cleanup is not necessary when in fact cleanup is necessary. Once these quantities have been specified we can determine (i) the number of samples n required in order to meet these specifications, and (ii) the rejection number r. If r or more of the n samples have activities greater than L, then cleanup is required. Note that this approach assumes we are willing to tolerate a certain proportion (p_1) of samples with activities greater than L without cleaning up the area. Of course, p_1 can be specified to be as small as we choose.

The risk β should be specified as a small quantity since the consequences of not cleaning up a contaminated area could be considerable to the inhabitants of the area. $1-\beta$ is known as the "power" of the design, i.e. the probability that the area is cleaned up when the actual proportion is p_2. On the other hand we would also like α to be near zero so as to avoid unnecessary cleanup operations. In the following table we give values of n and r for various values of p_1, p_2, α, and β. These were obtained using Table 13 in Burstein, H., 1971. Attribute Sampling; Tables and Explanations, McGraw-Hill, 464 pp. These values of n and r are for a non-sequential sampling plan. A sequential plan would probably require fewer samples.

From the results in TABLE 1 we note that:

a) As α gets larger the number of samples (n) required decreases when p_1, p_2, and β remain constant. Hence, if we are willing to risk spending more money on cleanup, the number of samples we need to collect decreases.

b) As β increases (power decreases) the number of samples n also decreases when p_1, p_2, and α remain constant. Hence, if we are willing to take a higher risk of missing some areas needing cleanup, we won't need to take as many samples.

c) As p_2 increases, the number of samples (n) decreases. If our cleanup criterion is that 10% rather than 2% of the samples must be greater than L before cleanup is started, then only 113 rather than 3063 samples need be taken (assuming $p_1 = \alpha = \beta = .01$). That is it will take many fewer samples to detect a difference between $p_1 = .01$ and $p_2 = .10$ than to detect a difference between $p_1 = .01$ and $p_2 = .02$. Hence, as p_1 and p_2 are placed closer together (for given α and β), the number of samples (n) increases.

TABLE 1

Number of Samples (n) and Rejection Numbers (r) for Nonsequential Acceptance Sampling for Specified Parameters α, β, p_1, and p_2.

	p_1 = .001				.01				.05			
p_2 =	.01		.10		.02		.10		.06		.10	
β	n	r	n	r	n	r	n	r	n	r	n	r
α = .01												
.01	1157	5	64	2	3063	45	113	5	10962	601	589	43
.05	773	4	46	2	2179	34	76	4	8091	451	448	35
.10	667	4	38	2	1782	29	52	3	7101	401	335	27
α = .05												
.01	838	3	44	1	2263	31	81	3	8339	451	435	30
.05	628	3	29	1	1567	23	61	3	5487	301	287	21
.10	388	2	22	1	1235	19	38	2	4515	251	222	17
α = .10												
.01	661	2	44	1	1939	26	64	2	6578	351	362	24
.05	473	2	29	1	1258	18	46	2	4614	251	227	16
.10	388	2	22	1	993	15	38	2	3647	201	175	13

The proper use of "Acceptance Sampling" requires that samples be collected at random within homogeneous areas (see, e.g., Sampling Inspection (H. A. Freeman, M. Friedman, F. Mosteller, and W. A. Wallis, eds.), Mc-Graw Hill, 1948, pages 48, 49 and 89). Concerning the homogeneity assumption, it seems advisable to divide an island into two, three, or more areas depending on general level of activity and to go through the acceptance sampling procedure in each area separately. These areas could be defined on the basis of the plutonium concentrations obtained by the 1972 survey.

The assumption of random sampling within areas is important in order to preserve the α and β risks decided on for the decision making process. The use of alternative sampling plans, such as sampling at grid nodes of a systematic grid, must be carefully evaluated and supervised to insure the integrity of the final decision. This is a most important consideration in the design of the cleanup study that requires attention to detail. Someone familiar with the statistical requirements should be in the field during the sampling process to insure fidelity to the agreed upon design.

We note that attribute sampling is ordinarily used in situations where the "attribute" can be measured accurately for each element examined and decisions about a given population (often a quantity of manufactured product) are to be made on the basis of the sampled elements. Hence we are neglecting "counter error" here and assuming decisions are to be made on the basis of whether or not sample elements from a given area (e.g., soil aliquots) indicate that a proportion of such elements are above some set limit.

Question 4: For certification of islands for which cleanup of Pu has been performed:

a) What data are required?
b) How are the data to be evaluated?
c) What goals that are likely to be attainable in terms of the assurance that can be given that the cleanup criteria have been met?

In Question 3 we suggested acceptance sampling as a method to decide whether cleanup is necessary. Following the cleanup operation additional soil samples and In-Situ measurements must be taken for certification. Acceptance sampling as outlined above could also be used for this purpose (see TABLE 1 for number of samples required). If the certification requirement states that all collected samples must have plutonium concentrations below the critical level L, then the values of n in TABLE 2 below are appropriate (calculated using Table 12 in Burstein). If any sample has activity greater than L then the cleanup operation has not been successful and certification would not be issued. β and p_2 are defined as above in our discussion of Question 3. Note that the α risk (of concluding that cleanup is necessary when it really isn't) is not specified in TABLE 2. This risk does exist, but is ignored here on the basis that risk β (of concluding that further cleanup is not necessary when it really is necessary) is the most crucial for certification purposes. Sampling for certification should also be done independently for homogeneous areas within islands.

TABLE 2

Number of Samples* Required to be 100(1-β)% Sure that the True Proportion of Samples With Concentrations Greater Than L is Less Than p_2

	p_2		
β	.01	.05	.10
.01	458	90	44
.05	298	58	29
.10	229	45	22
.20	160	31	16

---

*Based on assumption that we will find no samples with activities greater than L.

Question 5: For cleanup operations, is there some optimum combination of In-Situ, soil sampling, and wet chemistry measurements that yields the most relevant information to guide contaminated soil removal at the least cost? Can a generalized approach be developed for use with all islands or should guidance be derived for the known conditions on each island requiring change?

The question of optimum combination of In-Situ and soil sampling needs to be addressed relative to the kriging procedure. Hence, Dr. Delfiner should be consulted on this matter. In general the optimum combination will depend in part on how well the In-Situ and plutonium concentrations from soil samples are correlated, and on the relative costs of the two procedures. Gilbert and Eberhardt (1976, "An Evaluation of Double Sampling for Estimating Plutonium Inventory in Soil", Radioecology and Energy Resources, Proceedings of the Fourth National Symposium on Radioecology, Dowden, Hutchinson and Ross, Inc.) discuss the issues involved.

The question of a generalized approach should also be taken up with Dr. Delfiner. The general level and heterogeneity of plutonium activity in soil over an island will certainly affect the total number of samples required for cleanup (if any) and certification. However, the general sampling design may be applicable to all islands.

I hope this letter will help you in planning for the Enewetak sampling and cleanup effort. Some of the ideas discussed here are in pretty rough form and would need considerable thought to develop a final plan. Hopefully my brief comments on "acceptance sampling" will serve to stimulate discussion on its merits relative to the "average concentration" approach for deciding whether cleanup is required or has been achieved.

Best regards,

Dick

Richard O. Gilbert
Senior Research Scientist
Systems Department
Statistics Section

cc: Roger Ray, ERDA, NV, Las Vegas
Bruce Church, ERDA, NV, Las Vegas
Paul Dunaway, ERDA, NV, Las Vegas
Mary White, ERDA, NV, Las Vegas

Supplement to Letter from R. O. Gilbert to T. McCraw dated September 22, 1976 Concerning Sampling Plans for Enewetak Cleanup Survey.

I. Confidence Limits on True Average (Median) Concentration.

x = Pu concentration

$y = \log_e x$

If x is distributed lognormally, then

$\text{Prob}[\mu \leq \bar{y} + \frac{ts}{\sqrt{n}}] = 1-\alpha$ (since the y_i are normal),

where s = standard deviation of the y's.

$\bar{y}$ = mean of logs of the sample data,

μ = true (unknown) mean of logs

t = "t" value for specified α and n-1 degrees of freedom.

Then $\exp(\bar{y} + t_\alpha s/\sqrt{n})$ is an approximate $(1-\alpha)\%$ upper limit on the median of the lognormal distribution (original data). The median is that concentration above which and below which half the observations lie.

For Janet (data taken from Fig. B.8.1.i in NVO-140) we have

$n = 139$, $\bar{y} = 2.180$, and $s = 1.152$

For $\alpha = 0.01$, 0.05, and 0.10 we find:

α	t_{138}	100 $(1-\alpha)\%$ Upper Limit on Median
.01	2.35	11 pCi/g
.05	1.66	10
.10	1.29	10

Interpretation: For $\alpha = .01$ we state: We are 99% sure that the true (unknown) median Pu concentration on Janet is less than or equal to 11 pCi/g (if the data are lognormal).

Discussion: An alternative approach would be to assume the mean $\bar{x}$ of the Pu concentrations is approximately normally distributed. Then an upper confidence limit on the true (inknown) mean would be computed as $\bar{x} + \frac{ts}{\sqrt{n}}$, where s now refers to the standard deviation of the original untransformed observations. Since for Janet we have $n = 139$, $\bar{x} = 15.9$ pCi/g, $s = 20.9$ pCi/g we find the approximate limits:

α	t_{138}	100(1-α)% Upper Limit on True Mean
.01	2.35	20 pCi/g
.05	1.66	19
.10	1.29	18

Since the decision to cleanup may be a function more of extreme values rather than average concentrations the next section considers upper limits on percentiles.

II. Nonparametric Confidence Limits on Percentiles

Using "Practical Nonparametric Statistics" by W. J. Conover, John Wiley, 1971, page 111, we compute upper one-sided confidence limits:

> The probability is 1-α that p percent of the soil concentrations for the area from which samples were collected are less than or equal to X.

Estimated values of X for various values of p and α for the data from Janet are:

p	α	X(pCi/g)	
.50[†]	.01	13	
.50[†]	.05	11	(median = 9.8 pCi/g)
.50[†]	.10	11	
.50	.25	10	
.90	.01	51	
.90	.05	46	(90th percentile = 37 pCi/g)
.90	.10	41	
.90	.25	41	
.95	.01	120	
.95	.05	67	(95th percentile = 46 pCi/g)
.95	.10	57	
.95	.25	52	

Interpretation: For p = .90 and α = .05 we state: We are 95% sure that 90% of the soil concentrations on the island are $\leq$ 46 pCi/gm.

[†] These values of X for α = .01, .05, and .10 when p = .50 are nonparametric equivalents of the 100(1-α)% upper limits on the median computed in Part I above. The upper limits (X) obtained here do not require any assumption about the distribution of the observations. Note that these limits are consequently somewhat higher than the corresponding limits in Part I.

Note: These computations assume the data are homogeneous, i.e. there are no trends in the data. Since there <u>are</u> trends present on Janet (increasing concentrations near GZ areas) these kinds of computations should be done separately for GZ and low level areas.

III. One-Sided Confidence Limit on a Proportion

Using "Attribute Sampling" by Herman Burstein, Mc-Graw-Hill, 1971, (Table 1) we can obtain the following probability statement:

> The probability is $100(1-\alpha)$ that the proportion of soil samples with Pu concentrations greater than or equal to the cleanup Level L is less than or equal to P.

Estimates of P for various values of α for cleanup level 40pCi/g (using the 139 soil samples (0-15 cm) from Janet) are:

α	P
.01	.167
.05	.145
.10	.133

Note: Proportion of samples with Pu concentrations $\geq$ 40 pCi/g is 13/139 = .0935.

<u>Interpretation</u>: For $\alpha = .01$;
We are 99% sure that 16.7% of the soil samples on Janet have concentrations $\geq$ 40pCi/g.

<u>Discussion</u>: A possible approach to deciding whether an island needs to be cleaned up is as follows: The island (or parts of the island) will be cleaned up unless P is less than, say, 5% for some specified α level, say .01. If it had happened that only 1 of the 139 samples had a Pu concentration $\geq$ 40pCi/g then we find that $P = .047$ (4.7%) for $\alpha = .01$. Hence, in that hypothetical case we would decide not to cleanup the island if the above rule ($P \leq .05$ when $\alpha = .01$) had been used. An alternative and perhaps preferable method of deciding whether cleanup is necessary is discussed under Question 3, part B.