Battelle
Pacific Northwest Laboratories
Battelle Boulevard
Richland, Washington 99352
Telephone (509)
Telex 36921

August 28, 1972

409771

Dr. Don Hendricks
Director of Radiological Operations Division
Nevada Operations Office
USAEC
Box 14100
Las Vegas, Nevada 89114

Dear Don:

This letter is to summarize our understanding of the sampling aspects of the forthcoming Eniwetok survey. I assume there will be some further discussion of this, and this letter is intended to serve as a basis for such discussions. I have had several long talks with Oliver Lynch by telephone, and Dick Gilbert spent a day with him last week. We very much appreciate his cooperative spirit and have been much impressed with his understanding of the problems and thorough-going approach to the survey. Probably Ollie will not agree with all of our suggestions, since he has to implement them in the field. Having had occasion to modify my own survey designs in the field, I am aware of the kinds of things that come up. However, we do need to try to spell out some of the consequences of particular modifications where this is possible. Mostly these have to do with what random sampling amounts to, and the effects of deviations from a truly random sample. There is ample evidence that a representative set of sampling locations cannot be selected by "eyeball". There are quite a few practical-minded people who don't believe that statement, but we've got enough to do at the moment that I don't want to spend any time on that old argument. Nor can we say exactly what the consequences of departures from random sampling schemes will be. We will suggest some possibilities, and perhaps it is sufficient to note that if the results of this study have to be somehow defended in "public" they had better be based on a defensible scheme.

The balance of this letter will deal with some specific topics. Some of these we haven't looked into in the detail we feel advisable, but time is pressing on. In what follows we assume that there are reasons for wanting to make specific statements either about specific parts of an island, a single island, or about groups of islands. Our purpose is to try to provide some

criteria for determining the number of samples to be assigned to a particular such unit. We assume that all concerned understand that there is no single "best" sample size and that someone has to judge what criteria will serve the objectives of the survey--this process ought not to devolve on us, since we're not sufficiently familiar with the problem and the Atoll. A common tendency is to try to find an acceptable number of samples per unit area (say one per so many thousand square feet). This comes from the usual lay opinion that a big area should have more samples than a small one. Although a small adjustment can be made for size of area, it is generally true that the controlling factor is not size, but is variability from spot to spot. Sometimes this is a hard point for field workers to accept, but I think that most of those who have had a look at data on radionuclide concentrations due to fallout will remember that samples taken a few feet apart are often just as different as are those separated by much greater distances (see for example, P. 6 of the study by Held et al. of atoll soils and radionuclides; UWFL-92, 1965).

A good deal of experience with radionuclides in a variety of substances suggests that the coefficient of variation is relatively constant, if the data are taken from situations having some degree of homogeneity. This observation, plus other information, indicates that the frequency distributions of observed data will be "skewed", i.e., if one plots the relative numbers of soil samples having specific concentrations, there will be a sizable number at some point to the right of zero (with the appropriate scale) and a long "tail" off to the right. The commonly used distributions for representing such data are the lognormal and gamma distributions. The existence of such skewed distributions has some important implications when it comes to clean-up, and we will return to this point later.

Criteria for selecting sample sizes

If it is assumed that the coefficient of variation is a constant, i.e., that the standard deviation (s) divided by the mean ($\bar{x}$) is approximately constant, then we can compute a standard error of the mean:

$$\text{coef. of variation} = c = \frac{s}{\bar{x}} \quad \text{Standard error} = \frac{s}{\sqrt{n}} = \frac{c\,\bar{x}}{\sqrt{n}}$$

and express this as a proportion:

$$\text{Standard error as proportion} = \frac{s}{\sqrt{n\bar{x}}} = \frac{c\bar{x}}{\sqrt{n\bar{x}}} = \frac{c}{\sqrt{n}}$$

so that approximate confidence limits can be expressed as (for accuracy one might use t-variate as multiplier for the smaller sample sizes, but the present procedure seems adequate for making judgements; also we will later bring in the effect of skewness on confidence limits):

$$\text{Confidence limits as proportion of mean} \doteq \pm 2\ SE(\text{prop.}) = \pm \frac{2c}{\sqrt{n}}$$

Thus if we know or assume the value of the coefficient of variation appropriate for the Atoll, one can calculate the sample sizes necessary for statements such as the following:

"The probability that the true mean concentration of plutonium in soil samples from Japtan Island will be within $\pm$ 5% of the observed mean determined from n samples is .95". In other words, we can specify a percentage interval within which we can have high confidence that the actual concentration will fall, given an advance estimate of C (the coefficient of variation). Note that all we need is C, since the actual concentrations do not enter the calculations -- we make the statements as proportions or percentages.

Some actual estimates of C for soil plutonium are as follows:

Locations	Sample Size	Coefficient of variation
Bikini (page 12 of NVO-97/ Summary report of the 1969 and 1970 Bikini Surveys)	11	.74
Eniwetok (Data on samples from Janet, by phone from O. Lynch)	12	.42
Eniwetok (9 composites of 3 each from Phase IIA islands; Alice to Edna)	9	.75 (1.30)*
Rocky Flats (HASL-235)	33	3.6
Nevada Test Site (GMX study)		
— Stratum I	10	.66
Stratum II	4	.55
Stratum IV	6	1.36

*adjusting for composites

We have arbitrarily selected C=.7 as the value to use here. Other evidence suggests that this will not be a bad guide for other radioisotopes, and that the results should apply reasonably well to those islands contaminated mainly by fallout. If the sampling is done along a very distinct gradient in concentrations, then the above guides don't hold, as for example as shown in the Rocky Flats data. If one has some evidence on which to sketch in several levels of contamination (as was available at the GMX site on NTS) then the variability within such a subdivision should again be roughly as indicated. The consequences for the Eniwetok study seem to be that we can feel reasonably secure in predicting confidence limits (as %) for situations other than on those islands where ground zeros existed, where rather larger samples are indicated for comparable confidence limits.

The above calculations can be summarized in the following table:

n (sample size)	Approximate confidence limits as % of mean concentration
10	± 44%
20	± 31%
30	± 26%
40	± 22%
50	± 20%
60	± 18%
100	± 14%
200	± 10%

These results are simply calculated as $1.4/\sqrt{n}$ so one can easily find values for other sample sizes, or can increase the numerator (which is 2(.7) or twice the coefficient of variation) to some larger value to reflect supposedly greater variation. From this evidence, we don't recommend that sample sizes as small as 10 be regarded as anything but rough guides to relative concentrations. Samples of 30 or more may be sufficient to make some reasonably satisfactory estimates. However, if one wants to be fairly precise, even 100 samples is not too many, since we still have a fairly wide range (±14%) for the true mean. Perhaps it is worth repeating here that these confidence

limits are statements that can be made in advance of sampling or about many future repetitions of the same survey. The actual concentration of plutonium present is some particular value and is not likely to be changed by our confidence limits! It is important to note that these calculations say nothing about area or number of islands involved. As noted above, it is the variability, not the area, that determines sample size. Hence the size of sample needed goes up or down depending on how islands are grouped or subdivided and on what choices are made as to precision of results needed for a particular sampling unit.

A different criterion is available if one chooses to consider the size of sample required to determine whether concentrations differ from place to place. Presumably this sort of judgement may be involved in deciding or confirming whether the seaward and lagoon sides of various islands have different concentrations of plutonium, or whether the several sets of "fallout" contaminated islands have different concentrations. The kind of statements that can be made for given samples sizes now become somewhat more complicated, and there are other features having to do with number of places being compared and differences in sample size. For simplicity we present a comparison of just two areas (or islands or two groups of islands) and assume that each has been sampled with exactly the same number of plots. Then the comparison may be a "t-test" and we are interested in what the statistician calls the "power" of the test--i.e., what is the probability that we will actually detect a difference of a specified magnitude.? The results can be phrased in statements like "the probability is .90 that we will be able to detect a difference of 40% in plutonium concentration between seaward and lagoon sides of an island if we take n samples on each side of the island" (for purists, we note that one also needs to set the probability of Type I error, i.e., the chance that we claim a difference when none exists; we have here used the .05 level as is usual). Since the distributions are skewed we have assumed a logarithmic transformation and express the comparisons as a ratio (R), that is, we suppose the "high" area is R times the "low" area. A quick look at the table attached willshow that samples of even 50 won't distinguish (safely) anything but large differences, whereas with samples of size 30 (in each location) about 70% of the time one would expect to distinguish a difference as large as 50% (R=1.5). Larger differences will of course be picked up with

higher probability and larger samples will detect smaller errors. As a working guide, we recommend that a power of .90 be the mininimal value used to select sample sizes. Thus to select the sample size needed to make comparisons between two islands (two areas, two groups of islands) first determine what difference should be detected (e.g., 1.3 means roughly 30% differences will be picked up) and choose a sample size large enough to give a power of .90 or better (about 130 in this case). Lacking any firm choice as to criterion as to difference to be detected, one can simply browse around in the table and get a general impression of what difference a given sample size will serve to detect.

R (Ratio of "high" to "low" mean)

1.3		1.4		1.5		1.8		2.0		2.5		3.0	
n	β	n	β	n	β	n	β	n	β	n	β	n	β
100	.83	50	.75	30	.69	22	.86	14	.80	9	.83	4	.54
120	.89	75	.89	47	.88	25	.89	16	.85	11	.89	6	.76
130	.91	85	.92	55	.91	27	.91	20	.91	12	.92	8	.89
140	.93	100	.96	60	.93	30	.94	22	.95	13	.94	9	.93

In the above table, β denotes the probability that samples of size n will detect ratios of mean concentrations of size R. As an example, if we expect the concentrations in one area to be 1.3 times that in another and take 100 samples in each area, then the t-test will have a probability of .83 of finding this difference (and probability of .17 of not identifying this difference, even though one as great as this actually exists).

An inspection of the above table will show that one cannot have a high probability of finding differences of less than about two-fold unless rather large samples are used.

Sampling patterns

In beginning this letter, we pointed out the necessity for using random sampling if one is to be protected against criticism. There is one rather widely used alternative scheme, which is to locate samples in some systematic pattern. In a problem like the present one, a two-dimensional grid may be the most likely pattern, so that the samples are located at, say, x-hundred yard spacing. One argument for using such a pattern is that it may provide somewhat better estimates of the amount of soil to be moved in a clean-up

operation. I rather suspect that this is true only if the source of contamination is such that the concentrations can be regarded as having a smooth gradient in both directions from some maximum -- one might thus visualize a three dimensional plot of concentrations as a hill of some regular shape, so that evenly spaced sampling locations make plotting contours somewhat easier and efficient. However, in view of the history of contamination at Eniwetok I should doubt that there are many spots where this kind of pattern can be expected. Instead one might expect that the pattern would be sufficiently heterogenous that a "grid" pattern would give results that look much like a random pattern. Thus the main advantage to a systematic pattern would be ease of locating plots. This can be a distinct advantage, but there are also some further disadvantages. One that concerns me is the aforementioned evidence that the frequency distribution will be skewed. I suspect that it will be important to have the best data we can get as to the shape of that distribution and such information ought to come from a random set of sample points. The shape of the distribution becomes important, it seems to me, if it becomes necessary to make exact statements about the fraction of the area (or volume) that will exceed a specified concentration. That is, suppose it is decided that concentrations of more than x picocuries per gram must be "cleaned up". One then would like to be able to fit a theoretical curve and use it to make probability statements about the fraction of the area (volume) that exceeds x picocuries per gram. How one actually goes about that calls for some more discussion, a topic we'll return to below. For the present, I only want to make the point that random sampling is indicated.

The actual process of drawing a random sample has been discussed with Ollie Lynch. We recommend that a rather fine grid, at say, 50 foot intervals be laid down on a map and numbered. The actual locations should then be selected from a table of random numbers and plotted on the map. The order in which the sample points are located should be recorded, so that any last minute decisions to change sample size can be easily handled (i.e., by dropping the last samples selected; or by adding more if needed). The grid interval can be selected on practical grounds--i.e., how accurately it is assumed that a particular location can be found by pacing. However, it is wise to use a finer grid than such considerations may indicate, and it is essential to leave no ambiguity in instructions for field work--that is field crews should have exact instructions as to distances and directions and be told to sample at the

specified spots. It should also be understood that there is no purpose in replacing or substituting sample locations. If there is some obstacle to sampling at a specified location (e.g., a concrete pad) then that fact should be recorded in the field and no sample taken. There are two grounds for this point. One, the heterogeniety of concentrations is such that shifting over a few feet prohibits calling the new location typical of the old, and, two, that obstacles to sampling are important in clean-up considerations -- a chunk of concrete or a tree will offer problems, and the frequency of such problems ought to be estimated from the survey. While it is sometimes hard for field workers to accept the policy of "no substitution", I think it is an important consideration and ought to be a hard and fast rule. If it seems desirable, some extra allowance in sample size can be made to get "enough" actual samples, but the statistical evidence is simply "no sample point = no information" (or information of another kind, i.e., x% of the area has obstacles of a specified kind). If there is reason to treat certain kinds of areas differently, then where possible, those areas ought to be set up in advance as specified sampling units. In particular, disturbed areas might fall into such a class.

The above procedure may need to be cleared up in telephone conversations or other discussions. The essentials are to decide in advance what areas are to be treated as units, to set up maps of these units and to locate sample points at random on those maps. Where several islands are to be combined, the sampling should be at random over the group of islands, and not limited to a fixed number per island. If, however, a decision is made to fix the number of plots per island, we can handle this in the subsequent analysis. You will pay a price for this in the sense of taking a somewhat larger sample than needed for statements about islands as a group (or by getting a wider confidence interval).

Double-sampling

The above proceeds as though wet-chemistry analyses for plutonium are to be made for all samples. Very likely this is not practical, and we are assuming that it may be possible to use a combination of wet chemistries and readings by a Ge/Li ("Gelly") detector. To do this within the limitations of statistically acceptable practice calls for some finesse, and constitutes the biggest source of worry for us in the entire survey. For one thing, there are some technical points we're not quite sure of, and these may take a fair bit of time and effort to sort out. Obviously there may be places where the Am/Pu ratio may be so variable as to make the technique unusable.

Probably you can roughly identify those places in advance and proceed accordingly (i.e., more wet chemistries). Beyond that, the procedure should be as follows: (1) Determine what areas are to be treated as single sampling units (groups of islands, etc.). (2) Take a large random sample of soil plots and collect the soils. (3) Determine Am concentrations by Ge/Li detector on all samples. (4) Take a random subsample of all soil samples for the specified area and do wet chemistries on these. Our present thinking is that this sample should be at least 30 wet chemistries for each area. That is, if exact statements are to be made for some group of islands, or large island (or single area), these should each be based on at least 30 wet chemistries. I should expect that the southern islands (Phase I) could very likely be combined for one set of wet chemistries. However, if the soil samples are all kept and are available, then it should be possible to provide some modifications of the scheme as the results of the first set of wet chemistries become available--i.e., we start out with one set of analyses and do more as study of the data indicates. Ollie provided us with some advance data on the Phase IIA islands that indicates that the method should work fairly well there (high correlation between Am and Pu).

Our recommendation is to take rather more samples for the "Gelly" scans than the above criteria may indicate, since we have based these rules roughly on wet chemistries -- adding the double-sampling scheme will increase the variability. However, it should cut processing costs a good deal, and it seems prudent to take more samples than we think are needed and discard them if it turns out we don't need them--its much easier to collect them in the first pass than to send crews back again!

Additional points

The question of finding "burial grounds" by sampling has been mentioned. I should think that about all we can say in general is that one could guess as to the size of a burial ground (width of a bulldozer blade, guess as to length) and calculate that finding one by blind search is about equivalent to the needle-in-a-haystack problem. No doubt there may be some surface clues, etc., which one could use to devise a scheme, but more information and some

trials are needed if this is essential.

We are not very clear as to how to handle the problem of allocating soil profiles. In "fallout" situations, one mostly wants to confirm that the concentration drops off exponentially, so it may be sufficient to take some arbitrary number of the random locations and do profiles there. If the profiles are to be used in probability statements or in significance tests, they should be randomly spread through the areas of interest. Where profiles are to figure in clean-up criteria, we probably will need to try to design some specific plan. At present, we assume enough will be taken to provide an initial notion of the situation for various islands and groups of islands. Obviously double-sampling should be useful in this case. Again, we might take more than seem needed and decide on wet chemistries on the basis of an initial set of analyses.

Thinking ahead to the prospect of an actual clean-up operation, it seems to me that some experiments with compositing soil samples ought to be considered. My arguments are as follows: I hope that it will be possible to set up the criteria on a probability basis. That this is necessary follows from the skewed nature of the frequency distributions. If these hold then there will always be some probability that any given criterion will be exceeded by a small (in some cases, vanishingly small) number of plots. One possible scheme is to subdivide the area into rectangles (or squares) of a size convenient for e.g., earth removal, and to take a sizable number of samples from each such block in turn, and to combine these into several composites and analyze the composites. If the mean of the composites exceeds a predetermined value, then the plot is slated for clean-up. Such a procedure, perhaps using double-sampling, might reduce the number of samples that have to be analyzed. The immediate problem is that we know very little about the behavior of the compositing operation, and I'd hesitate to actually propose such a scheme without some data as to how well compositing works in practice.

I hope that the above discussion will give you some background as to our notions of the statistical aspects. We recommend that the entire sampling scheme be put together as soon as possible. Numbering the samples in order of the draw will permit such changes as become necessary in terms of field conditions. If the whole scheme is laid out in advance, there should then

be a chance to review it unit by unit and suggest changes as needed to meet various people's judgement as to what is essential.

This letter has been written in haste, and may thus be ambiguous in places. Please don't hesitate to call for clarification.

Sincerely yours,

L. L. Eberhardt
Research Associate
Ecosystems

LLE:sac

cc: EM Douthett NVOO
P Dunaway NVOO
O Lynch NVOO
RO Gilbert BNW
V Schultz WSU