

**DISCLAIMER**

This document contains information that is classified as "Confidential" by the United States Government. It is the property of the United States Government and is loaned to you for your use only. It is not to be distributed outside your agency without the express written consent of the United States Government. The United States Government assumes no responsibility for the use or misuse of this information. Reference herein to any specific product, process, or service, by trade name, trademark, manufacturer, or otherwise, does not constitute an endorsement, recommendation, or approval by the United States Government of any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

SOME DEFICIENCIES IN ANALYZING LEACHATES  
AND REPORTING RESULTS\*

C. W. Sill

EG&G Idaho, Inc.

Idaho Falls, Idaho 83415

USA

**MASTER**

Prepared for submission to the ORNL Conference on  
Leachability of Radioactive Solids, Gatlinburg, Tennessee,  
December 9-12, 1980.

\*Work supported by the U. S. Department of Energy, Assistant  
Secretary for Nuclear Energy, Office of Waste Operation and  
Technology, under DOE Contract No. DE-AC07-76ID01570.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

## ABSTRACT

Determination of the leachability of elements from various waste forms for long-term retention of radioactive wastes requires chemical analysis of both the leachate and waste form for the various elements or radionuclides of interest. Neither the precision of the leaching procedure nor the accuracy of the results can be determined unambiguously if gross uncertainties exist in either the analytical methods used or the way in which the results are reported and/or interpreted. Perhaps the most important consideration is to demonstrate beforehand that all participating laboratories can obtain the same analytical result within some stated uncertainty when the same concentration is known to be present, i.e., by analysis of a single homogeneous solution. Special precautions must be taken to ensure that all material leached from the waste form is recovered from the leaching container and that all refractory particles are dissolved completely before analysis is begun. The actual results obtained should be reported, including negative signs if present, avoiding all subjective and quantitatively nondescriptive statements such as nil, none, not detected, etc. Each result must be accompanied by an estimate of its uncertainty which must include all random uncertainties incurred anywhere in the entire measurement process, preferably expressed as the standard deviation. Problems involved in interlaboratory comparisons are also discussed.

SOME DEFICIENCIES IN ANALYZING LEACHATES  
AND REPORTING RESULTS

INTRODUCTION

One of the most promising methods for the long-term disposal of radioactive wastes involves incorporation of the waste into glasses, basalts, or other extremely refractory materials from which the radionuclides can be leached only at very slow rates. To determine the most suitable composition for the waste forms, samples of each type must be leached under a variety of experimental conditions expected to be encountered during storage, and the quantity and type of each radionuclide removed is determined by analysis of the leachate. If the waste form is to be successful, the quantities of radionuclides removed, even over long periods of time at elevated temperatures, must be very small indeed. Thus, only analytical procedures of the highest sensitivity and precision will be suitable to measure the extremely low concentrations being sought. In many cases, determination of stable elements will also be desirable at comparable sensitivities to permit understanding the mechanism of the leaching process.

To intercompare results from different laboratories, a single standardized leaching procedure will be used by all participants to determine the leachability of a given waste form under a given inflexible set of conditions. The reproducibility of the procedure

will then be determined by comparison of participants' results. It seems self-evident that neither the precision of the leaching procedure nor the accuracy of the leaching rates can be determined unambiguously if there are gross uncertainties in the analytical methods used, or in the way in which the results are reported and interpreted. This paper will discuss some examples of both kinds and recommend methods for their correction.

#### THE STATE OF THE ANALYTICAL ART

Before the precision and accuracy of a leaching procedure can be determined, it must be demonstrated that the participating laboratories can all obtain the same analytical result on a given homogeneous solution within some stated uncertainty. Unfortunately, analytical accuracy cannot be accepted as either a justifiable or a foregone conclusion. Many different elements and radionuclides will be measured over a wide range of concentrations by analytical procedures that could conceivably range from the simplest chemical ones to those involving the ultimate in expensive and sophisticated instrumentation. Consequently, the sensitivity, precision, accuracy and reliability of the results will vary greatly, particularly when each laboratory is invited to use its own procedures without other instruction or guidance. Different methods can produce results differing by many orders of magnitude, and useful comparisons are not possible unless the analyses are carried out under strictly comparable conditions. When a large scatter among the results is obtained,

outlying results are often discarded, and the mean and experimental standard deviation of the remaining results are reported. In fact, the dispersion obtained is much more likely to reflect the imprecision and inaccuracy of the analytical measurements resulting from a diversity of procedures and analysts. The difficult part of the argument is that each laboratory can generally produce a highly reproducible set of numbers to support its claims of accuracy. However, when several laboratories analyze the same demonstrably homogeneous sample, the wide variety of results invariably obtained quickly proves that not all of the laboratories can be correct in their assertions.

A large quantity of such analytical data from interlaboratory comparisons around the world was presented recently (1) to demonstrate the generally unsatisfactory performance being achieved. For example, ten well-homogenized samples of soil were analyzed for Ra-226 by ten different laboratories that had been selected as being some of the most competent in the country in this particular determination (2). The individual results obtained by the participating laboratories on one of the samples were  $8.6 \pm 2.5$ ,  $1.64 \pm 0.45$ ,  $5.5 \pm 0.2$ ,  $2.48 \pm 0.05$ ,  $0.00 \pm 0.01$ ,  $1.05 \pm 0.28$ ,  $1.72 \pm 0.04$ ,  $2.0 \pm 0.3$ ,  $2.83 \pm 0.69$  and  $3.3 \pm 1.7$  for a mean and experimental standard deviation of  $2.91 \pm 2.48$  pCi/g. The dispersion of the results was almost as large as the mean itself. Part of the problem is caused by the wide variation in the precision with which the measurements were made. Obviously, the results themselves cannot be any more accurate

than the uncertainty in the measurements will permit. Although many of the results are in statistical agreement because of the large values of the standard deviations, there is no consolation to be derived from this fact in determining the "best value" for the sample. The results on the other nine samples were equally bad, showing that the example cited was not misleading. The means and experimental standard deviations of the ten results on each of the other nine samples were  $2.85 \pm 2.54$ ,  $5.31 \pm 2.39$ ,  $8.95 \pm 4.12$ ,  $3.21 \pm 2.58$ ,  $4.72 \pm 3.04$ ,  $3.00 \pm 1.52$ ,  $3.12 \pm 1.70$ ,  $7.38 \pm 4.28$  and  $11.59 \pm 3.04$  pCi/g. The number of significant figures retained are those of the compiler, not the present author. Unfortunately, samples of known concentrations were not included in the intercomparison so that there is no way of telling which results, if any, are correct. The only statement that can be made about these data is that at least some of the laboratories have a problem, both in the precision and accuracy of their results and in the way the data are reported. These problems will be discussed later.

In Italy, 14 laboratories from different agencies responsible for monitoring environmental radioactivity analyzed samples consisting of gross beta activity in simulated air filters, Sr-90 and Cs-137 in simulated dry deposits, and Co-60, Sr-90, and Cs-137 in surface water (3). Typical of the conclusions reached by those authors were "...most Sr-90 results are excessively biased, showing an unacceptable inaccuracy, while precision of the Sr-90 measurements was unacceptable for only two laboratories". Also, "The results of [7] laboratories

for Cs-137 measurements were excessively biased; ... precision of Cs-137 measurements was acceptable for all the laboratories". Generally, the precision in each laboratory was good to excellent but the accuracy was very poor. As has been observed many times, high internal precision and extremely poor accuracy seem to be more the rule than the exception.

There seems to be a general feeling that the small di- and trivalent elements are not as difficult to determine as the more hydrolytic and refractory ones such as the large ter- and quadrivalent actinides, and that the use of expensive and sophisticated instrumentation avoids all the problems of wet chemical analysis. That neither assumption is necessarily true is amply demonstrated by the results of an intercomparison study carried out recently (4). Thirty-five laboratories from 19 countries, including two top laboratories from the United States, participated in the determination of 16 trace elements in water. All participating laboratories had a well-established reputation in trace metal analysis. Atomic absorption spectrometry and neutron activation analysis were used to obtain about 80% of the results. The results were generally very poor as indicated by the authors' interesting observation: "Although the spread of results supplied by particular laboratories is high (in some instances exceeding two orders of magnitude), the over-all mean calculated after rejection of outliers usually approximates closely to the true value...." This profound statement suggests that all you have to do to get the correct answer is to have 35 laboratories from

19 countries analyze all your samples and take the mean. The same authors also state: "A wide spread of results in trace-element analysis seems to be typical at present. In some other intercomparison tests in which all participating laboratories had a well-established reputation in trace metal analysis, values differing by a factor larger than 10 and sometimes larger than 100 were obtained for some elements."

Many laboratories depend heavily on internal replication to evaluate the adequacy of their analytical capability. Even when excellent agreement among the results is obtained, replication can only indicate precision, not accuracy. Even intercomparison between laboratories is generally ineffective in determining by consensus the true concentration of a given element in a sample because of the variability in the results generally obtained. Philosophically, one should not feel particularly secure or confident when his result agrees with that of the majority, but neither should he feel necessarily inferior when his result qualifies as an outlier by any available statistical test. There are many cases in which it can be demonstrated conclusively that the outlier was the most nearly correct result in the group.

Interlaboratory comparisons are very expensive to carry out, and few laboratories have the time, the inclination, or the capability to contribute in any very meaningful way. The time and effort would be far better spent in developing and analyzing standards containing

exactly known quantities of any given radionuclide by which even the smallest laboratory could determine, unequivocally, quickly, and without assistance from others, how both its personnel and procedures were performing under routine conditions. Preparation of such solid standards has been discussed elsewhere (1). Despite claims to the contrary, solution standards for use in determining the accuracy of analyses of aqueous leachates can be prepared simply and reliably from primary standard compounds, using sufficient acid or complexing agents to prevent incipient hydrolysis, ion exchange, etc.

#### SAMPLE RECOVERY AND DISSOLUTION

An important analytical problem in obtaining accurate results is that of getting all refractory material removed from the waste form out of the leaching container so that it can be dissolved. This insoluble material must be completely converted to a true ionic solution before the chemistry characteristic of the element being sought can be expected to be displayed. It should be well known, particularly to chemists, that many ter- and quadrivalent ions can be precipitated from uncomplexed solutions as hydroxides at pH's greater than about 3 or 4. At the pH of about 7 at which most leaching tests are carried out, some ter- and quadrivalent elements will surely be deposited on the sides and bottom of the leaching container, particularly during leaching periods of several days at elevated temperatures. Special treatment with acids, complexing agents, ultrasonic baths, etc. is required to remove these materials from the

container completely so that they will be included in the analysis, and to homogenize the leachate before subdividing for whatever purpose. Otherwise, the results will be seriously in error, and severe contamination will result when the leaching container is reused. For example, a recent document specified that the leachate be measured and divided into two equal parts. One half was to be stored for future reference; the other half was to be returned to the leaching container, acidified with nitric acid and allowed to stand for 1 week before analysis. Apparently, the author was aware of the necessity of adding acid and allowing a substantial amount of time for the hydrolyzed material to redissolve. However, if hydrolyzed material were present, and part of it were poured out of the container, putting half of the leachate back into the leaching container with acid to dissolve the remaining material will certainly unbalance the leachate irretrievably. The reference half will obviously be completely different from the half on which the analyses are made, and neither of them will be correct.

While the strictly hydrolytic deposits or material deposited on the walls by various sorptive processes might be dissolved and reclaimed rapidly and completely by treatment with appropriate acids or complexing agents, refractory particles will certainly not be. If corrosive leaching occurs by dissolving the more soluble components such as silica and the alkali-metal oxides, leaving small refractory particles containing iron, aluminum, zirconium, etc. detached from the specimen, such particles will require more severe attack than casual

acidification, even with hydrofluoric acid, to achieve complete and reliable dissolution. This effect was clearly demonstrated during a recent investigation of the fractionation of radionuclides that occurs in liquid samples from nuclear power facilities (5). Using gamma-ray spectrometry, the authors studied the distribution of several fission and activation products among the filtrate, the filterable material, and the container walls. Even from strongly acid solutions not containing any visible particulate matter, certain radionuclides such as zinc, cobalt and manganese, which are not known for their ability to form refractory compounds, were found in large part either in the filterable material or on the walls of the container. Isotopes of the same element such as Co-57, -58 and -60 showed substantial differences in their distribution ratios. Isotopes of chromium, zirconium, niobium, ruthenium and cerium were present in the insoluble fraction but were not found in the filtrates of any sample. Clearly, refractory particles not related to but containing the radionuclides involved were responsible for the deviation from the solubilities normally expected. Generally, the particles retained on the container walls were extremely difficult to remove even by repeated treatment with the strongest acids, including hydrofluoric, and with vigorous shaking or treatment in an ultrasonic bath. Addition of paper pulp to the empty container before addition of the sample was effective in keeping most of the particles off the container walls. It seems entirely likely that similar problems will be encountered with refractory particles formed during corrosive leaching of waste forms. This question should be investigated.

Because of the obvious inability to carry out one of the powerful high-temperature methods of dissolution of refractory materials in a plastic container, there is no sure way of guaranteeing complete removal of nonradioactive elements from the leaching container. The evidence cited above (5) shows that the refractory particles are removed very slowly and incompletely by each of several successive treatments with the strongest acids and complexing agents. Hopefully, the compounds leached from the waste forms might not be quite as intractable. However, until more definite information can be obtained about the minimum requirements for complete removal of leached material from the container, the following procedure might be helpful. At the end of the leaching period, swirl the solution vigorously to remove as much as possible any sediment deposited on the waste form. This should not change the conditions of the leach test significantly. Remove the specimen and replace the basket in the leaching container. Rinse the waste form briefly with distilled water, discarding the rinse, and dry, weigh, and perform surface analyses as desired. Nondestructive tests such as pH or conductivity measurements can be made directly on the leachate in the original container if desired, provided care is exercised to rinse the electrodes with dilute acid and copious quantities of water to avoid cross contamination. Obviously, all washes are to be discarded and the electrodes dried before insertion back into the solution. No aliquots may be removed from the original container for any purpose before complete homogenization has been accomplished. Inaccurate results will be obtained and the solution unbalanced irretrievably if

the solution is in fact inhomogeneous from material precipitated from solution.

Add concentrated nitric acid in the amount of about 5% of the total volume of leachate to the original leaching container containing both the entire leachate and specimen basket. Place the container in an ultrasonic bath for 15 to 30 minutes to dissolve or dislodge as much as possible of any material adhering to the interior walls of the container or basket. Small aliquots can then be removed for analysis for some of the less refractory elements, if desired, keeping track of the total volumes removed.

The more refractory elements such as zirconium, thorium, plutonium, etc. will require more powerful methods to guarantee complete dissolution before accurate results can be obtained. Transfer the remaining leachate to an appropriate container for chemical dissolution, measuring the total volume in the process. If radiochemical analyses are desired, the leachate can be evaporated to a pyrosulfate fusion in a 250-mL Erlenmeyer flask. If stable elements are being determined, the solution might have to be evaporated in a platinum dish and treated extensively with hydrofluoric acid to obtain complete dissolution without increasing the background and salts to undesirable levels.

The effectiveness of recovering hydrolyzed material or refractory particles detached from the waste form by corrosive leaching must be

verified. Add 50 mL of boiling 1-to-1 nitric acid and 5 mL of 48% hydrofluoric acid to the empty leaching container containing the specimen basket, replace the cap and allow to stand overnight in a warm place. Decant the solution into a 100-mL platinum dish, evaporate the solution to dryness, and analyze for the nuclides of interest. If the quantity found is very large relative to that present in the leachate, additional or more vigorous treatment of the container will be necessary to ensure complete recovery. The most unequivocal way of determining the efficiency of removal would be to prepare the candidate waste form initially with gamma-emitting radioactive tracers of each of the elements of interest. The kind and quantity of each radionuclide remaining in the leaching container after each wash could then be determined quickly and unambiguously by direct gamma-ray spectrometry on the entire leaching container. The leaching rates of the actual radioactive species of interest could be determined at the same time.

#### ANALYTICAL PROCEDURES

Although the ultimate use of the waste form will be for storage or disposal of radioactive wastes, much of the current effort on leachability is being carried out using stable elements. This practice is understandable to obtain information on the leachability of the stable matrix elements to understand the process. It also takes advantage of the sensitivity, speed and convenience with which many such elements can be determined by modern instrumentation such as

flame or flameless atomic absorption (AA or FAA), inductively coupled plasma emission spectroscopy (ICP), etc. However, the intent of some investigators seems to be simply to avoid the use of radioactive tracers as long as possible. This reluctance not only ignores the many valuable benefits to be derived from tracer techniques but also leads to some extremely bad and indefensible practices. For example, one hears frequently about studies using uranium as a "stand-in" for plutonium. Why, because of its lower specific radioactivity? Although there are distinct chemical similarities between the two elements when both are in the same oxidation states, their oxidation potentials are so vastly different that leaching rates will certainly not be even remotely similar if the rates are very high. If one desires to make a statement about the leachability of plutonium, the leaching tests and analyses had better be made with plutonium. Anything else is inaccurate and unscientific.

There are excellent radiochemical analytical procedures employing alpha spectrometry that can be used to measure several of the transuranium elements simultaneously (6,7). Extremely high sensitivities can be achieved with quantities of tracer so small as to pose little health hazard, the results are completely unequivocal, and the distribution of several different nuclides can be determined simultaneously for each waste form prepared and each leaching test conducted. Even several stable matrix elements can be determined quickly and sensitively by preparing the waste form with gamma-emitting tracers of each of the elements of interest, and

determining them simultaneously by gamma-ray spectrometry. Particularly when concentrations in the range of parts-per-billion are to be measured, radiochemical tracer techniques generally result in lower detection limits and are more unequivocal than direct measurement of the stable elements whose backgrounds generally become substantial in this concentration range. For example, although the fluorometric and mass spectrometric determinations of uranium are among the most sensitive methods known, their applicability is limited by the natural uranium present in the background, reagents, etc. In contrast, spiking the waste form with U-232 tracer, which does not occur in appreciable quantities in the environment, can be arranged to give equal or better sensitivity, and the alpha energy spectrum makes the results completely unequivocal.

One of the most obvious deficiencies in many analytical procedures is the assumption that acidification, even with hydrofluoric acid, will necessarily dissolve whatever chemical forms might be present. When refractory oxides, silicates, carbides, etc. are present, high-temperature fusions with anhydrous potassium fluoride (6) or pyrosulfate are generally required to ensure complete dissolution. A big advantage of this combination is that complete dissolution of virtually all refractory compounds can be ensured by merely carrying out the fusions. Both fluoride and silica are removed by volatilization during the transposition of the fluoride cake with sulfuric acid. The biggest disadvantage is that the sample ends up in a sulfate system containing high concentrations of alkali metal

salts. This is not a problem when radiochemical analyses are being made but might be expected to result in larger blanks when stable elements are being determined. In either case, the entire analytical procedure must be developed as a whole to make its several parts compatible with one another. This has been accomplished for most of the alpha emitting radionuclides (6).

When such powerful procedures cannot be used because prohibitively high blanks or high concentrations of sulfates and alkali metals would result, other means must be used to ensure complete dissolution of refractory particles if accurate and reproducible results are to be obtained. This is true even when such powerful techniques are used as ICP with a plasma operating at temperatures above 5000°C. Even if such particles are vaporized sufficiently completely and rapidly in the plasma to be measured quantitatively, which itself needs to be proved, the sample might still require prior dissolution to ensure homogeneity, without which neither precision nor accuracy is possible. Prolonged treatment with hydrofluoric acid is perhaps the best choice when maintenance of low concentrations of salts and backgrounds of stable elements is an overriding consideration. However, treatment must be prolonged or low and erratic results will be obtained.

## REPORTING ANALYTICAL RESULTS

Analytical results are frequently reported and interpreted as though they were absolute invariant numbers. Quite the contrary, every measurement involves both random and systematic uncertainties, without knowledge of which the results themselves are either completely uninterpretable or are interpreted incorrectly. This is particularly true when the random uncertainties involved are as large as those in the measurement of radioactivity. For example, it is not possible to say with any certainty that 72 dpm/mL indicates a higher concentration than one of say 46 dpm/mL until the uncertainties in both measurements are considered. Consequently, even when the analytical measurements are carried out correctly, much of the effort goes for naught because of the way in which the results are reported. Most of us use either too little statistics to be able to evaluate the random uncertainties properly or so much as to give the impression of greater precision and accuracy than either the chemistry or the precision of measurement can possibly justify. We should keep in mind the admonition of T. C. Chamberlin, in "Science" in 1899: "The fascinating impressiveness of rigorous mathematical analysis, with its atmosphere of precision and elegance, should not blind us to the defects of the premises that condition the whole process." The following comments are made particularly with the determination of radionuclides in mind but are equally applicable after some modification to the determination of the stable elements.

### Report What You Get!

Every measured value should be reported as a finite number. Do not report analytical results as subjective conclusions in quantitatively nondescriptive terms such as nil, none, trace, nealigible, not detected, below detection limits, etc. Even zero is not quantitatively descriptive, except in the sense of the last indivisible particle, atom or count. Do not report results as being "less than" some numerical detection limit because of the difficulty of averaging large numbers of such results and the consequent loss of information. For the same reasons, do not round negative values to zero. Report what you got, including the negative sign. I am not suggesting that there is such a thing as a negative concentration. However, there certainly is such a thing as a negative measurement where random processes are involved. In fact, on samples containing no detectable activity, there is identically the same probability that the sample will count less than the background, and by the same amount, as that the sample will count higher. Consequently, over a large number of measurements there should be the same number of negative net values as positive net ones. If not, the procedure is not properly balanced, i.e., there are other blanks or backgrounds that have not been properly evaluated. To retain small positive numbers that are not statistically different from zero simply because they are positive, and to reject equally small negative numbers only because they are negative is statistically indefensible and will result in an eventual bias in the average.

### Report Uncertainties!

Every measured value of a random process has an uncertainty with which the measurement was made that must be recorded with each result, preferably expressed as the standard deviation. Without the uncertainty statement, the result itself cannot be interpreted correctly. For example, two numbers that are numerically different can in fact be statistically identical depending on the value of the standard deviation. This frequently causes expensive and time-consuming searches for the cause of the difference which was never real to begin with. Conversely, two numerically identical values can in fact be markedly different, again depending on the uncertainties involved. Although a numerical similarity between an analytical result and the known value might be comforting to some, the comparison might be completely misleading if the uncertainties involved show that the two values might also be statistically different by a large percentage. For example, the identical results  $21 \pm 6$  and  $21 \pm 6$  could be different by more than a factor of three at the 95% confidence level. This is one of the drawbacks to the use of unweighted means: it assumes that the numbers used are as represented, i.e., absolute and invariant.

### Ensure Decimal Agreement!

The number of decimal places retained in the measured value must agree logically with the number of places in its statistical

uncertainty. The standard deviation is the quantity that predicts in which decimal place there will be statistical significance so that arithmetic rules of significant figures do not apply. For example,  $1.2 \pm 0.004$  and  $1.234 \pm 0.4$  are both incorrect. Either the result is as precise as indicated by the standard deviation of  $\pm 0.004$  and three decimal places should be retained in the result, or the number is as imprecise as indicated by the  $\pm 0.4$  and the second and third decimal places should be dropped, i.e., either  $1.234 \pm 0.004$  or  $1.2 \pm 0.4$  depending on which is correct.

#### Propagate All Contributing Uncertainties!

The uncertainty of each measured value must include all significant random uncertainties incurred anywhere in the entire measurement process. Each uncertainty should be propagated to the final result using the partial different equation known as the Law of Propagation of Errors. Incidentally, this title is a misnomer and should refer to uncertainties rather than to errors. Without a doubt, many results being reported today are given an incorrect uncertainty based on only the counting statistics of the sample and instrumental background, and ignore equal or larger uncertainties resulting from reagent blanks, chemical yields, standardization of tracers and counters, counting times and efficiencies, sample sizes, etc. All conceivable sources of random uncertainties must be evaluated and included in the error propagation if they will affect the uncertainty of the final result within the number of significant figures

justified. An estimate of the systematic error should also be stated if it is expected to be significant compared to the uncertainty of the measurement, e.g.,  $0.052 \pm 0.004$  ( $\pm 25\%$ ) pCi/g. Error propagation must also be extended to all values derived from the actual measurements such as ratios, percentages or differences of two numbers, material balances, sums and differences, etc. For example, a recent document contained tables showing fractions of Cs-137 leached from a waste form to five significant figures without uncertainties of any kind. If the uncertainties of the measurements had been propagated to their ratio, it would have been clear that the last two or three figures were insignificant and should not have been retained. Similarly, another document showed a value of 100.00% for the sum of analytical measurements of 30 different elements, a chemical impossibility!

### Be Realistic!

Do not permit any measured value to even have the appearance of precision or accuracy that it cannot possibly have! I think it is fair to say that many statisticians do not fully understand the limitations in the numbers with which they are asked to work. Like professionals in other disciplines, some statisticians become so engrossed in theoretical aspects that they forget some of the limitations imposed by practical application. However, most of the fault must lie with those responsible for the data. I am sure that many of the inconsistencies and disagreement in our analytical results

are caused by results that are made to look better statistically than they really are. If you must err, err on the side of throwing away some justifiable precision and accuracy rather than to imply too much.

Many scientists, including statisticians, report standard deviations to three or more significant figures. One of the more extreme examples was published just recently in a national journal. The means and standard errors of the Cs-137 content of ten different ecological samples were reported to six and five significant figures, respectively, on nine of the results, and to seven and six figures, respectively, on the tenth, i.e.,  $26404.21 \pm 1540.76$  pCi/g. Because only two out of three results are expected to lie within plus or minus the entire value of the standard deviation, because many other sources of uncertainty are frequently omitted from the standard deviation, and even when everything is included in the error propagation, the equations used are only approximations, why delude ourselves with the illusion that we can justifiably interpret a standard deviation to three or more significant figures? In the example given, would  $2.64 \pm 0.15 \times 10^4$  pCi/g really have changed the justifiable interpretation in any significant way? The usual excuse that "the computer did it" notwithstanding, it is clearly the responsibility of the originator of the data to ensure that his results give the correct impression before he releases them to others.

I like to see standard deviations on the final result be limited to one significant figure, or to two significant figures if the first

digit is a 1 to avoid large percentage differences. In other words, if the first digit in the standard deviation, without regard to the position of the decimal point, is a 1, retain the first two digits, e.g., 0.0016. If the first digit is larger than 1, round the number and retain only the first digit, e.g., 0.0035 becomes 0.004. The result itself is then rounded such that its last significant figure will be in the same decimal place as that of the uncertainty, as discussed above.

#### TREATMENT OF RESULTS FROM INTERLABORATORY COMPARISONS

Interpretation of results from interlaboratory comparisons is statistically difficult because of the small number of results that are generally obtained, and their poor agreement with each other. It is common practice to use the well known root-mean-square equation to calculate an experimental standard deviation for a group of results. However, this procedure only shows the dispersion of the individual results actually obtained about their mean, and ignores completely the precision with which the measurements were made. This is an extremely important point because few intercomparisons are ever carried out with the requirement that the precision of measurement be kept approximately the same for all laboratories. One would certainly not want to give the same weight to a measurement made with an uncertainty of 25% as that given to one made with an uncertainty of 1%. Furthermore, when the group of results is small, as is frequently the case with complicated analyses, chance becomes a very important factor

i: determining whether or not either the same mean or the dispersion would be obtained if the measurements were repeated. For example, during an intercomparison on the determination of plutonium in soil, one laboratory reported results of 35.1, 35.1 and 34.6, giving a mean and experimental standard deviation of the individual results of  $34.9 \pm 0.3$  dpm/g. What do you think of such precision? Unquestionably excellent! If I tell you that the known value was  $34.8 \pm 0.1$  dpm/g, what would you think of the accuracy? Absolutely phenomenal! Now suppose that I tell you that the statistical uncertainty in each of the three individual measurements was  $\pm 3.4$ . Now what would you think with respect to both accuracy and precision? Extremely lucky! Such excellent agreement among such imprecise measurements is highly improbable but it happened! In fact, it has been my experience that the improbable happens more often than most of us think. If identical values of 35.1 had been obtained and the counting statistics were poor, the experimental standard deviation of  $\pm 0.0$  would call attention to its almost certainly having been largely a matter of luck. However, if the standard deviation is almost any finite number, it is generally accepted because it is "what I got."

To avoid the illusion that your results are more precise than the randomness of the disintegration process itself will allow, the experimental mean and standard deviation should be compared with a weighted mean and standard deviation calculated from the precision with which the individual measurements were made. If the experimental deviation is the larger, an experimental error is indicated. This is

both understandable and expected. However, if the experimental uncertainty is smaller, indicating that you are doing better than the limiting uncertainties allow, an element of luck is indicated and the safest course would be to use the theoretical deviation. The main purpose of interlaboratory comparisons is to attempt to get at the accuracy of the measurements by using as many different procedures and analysts as possible. This weighting procedure gives the most weight to those results that are the most precise but which are not necessarily the most accurate, thus defeating the purpose of the intercomparison. On the other hand, there is no statistical process that can select from a group of numbers those that are accurate and discard those that are not. It is of the most fundamental importance to interlaboratory comparisons that all participating laboratories demonstrate that their procedures and personnel are accurate before any sense can be made of their results by any statistical process. Once this difficult task has been accomplished, the results can then be weighted according to the precision with which the measurements were made. However, this option is still not entirely acceptable by itself, as will be discussed below.

The recommendations for calculating weighted means and standard deviations being made by certain laboratories can give highly incorrect results. Some of the laboratories involved have excellent reputations, large followings, and high visibility. Even some of the statistics books are not completely clear in this regard. The deficiency arises from the use of the reciprocal of the variance as

the weighting factor. When all measurements are made with approximately the same precision, as in most replicate measurements made within a single laboratory, the method gives acceptably correct results. However, when the measurements vary widely in precision, means and standard deviations weighted in such a manner can give results that are completely unreasonable. This is particularly true in weighting results of interlaboratory comparisons in which each laboratory is invited to use its own procedures and both the result and its uncertainty vary greatly. The problem arises from the fact that this method cannot distinguish between an absolute standard deviation that is small because the measurement was made in a very precise manner and one that is small simply because the result itself was very small. Consequently, the smaller the absolute value of the standard deviation becomes for whatever reason, the greater the weight given to that result (as the square) without regard to the precision with which the measurement was made.

In my opinion the weighting factor should be the reciprocal of the square of the relative standard deviation rather than of the absolute value. Thus,

$$W_i = \frac{1}{\left(\frac{S_i}{\bar{X}_i}\right)^2} = \left(\frac{\bar{X}_i}{S_i}\right)^2, \text{ where } W_i \text{ is the individual}$$

weighting factor,  $X_i$  is the individual result, and  $S_i$  is the estimate of the standard deviation for the individual result. The latter form should be used so that the calculation will not require division by zero when the result is zero. We can confidently expect that there will not be many cases in which the estimate of the standard deviation will be zero. Because the weighted standard deviation will now come out in relative terms, it must be multiplied by the weighted mean to obtain the absolute value. Thus, when

$$W_i = \left(\frac{X_i}{S_i}\right)^2, \quad \bar{X}_w = \frac{\sum W_i X_i}{\sum W_i}, \quad S_{rel} = \left(\frac{1}{\sum W_i}\right)^{1/2} \text{ and } S_{abs} = \bar{X}_w S_{rel}.$$

To demonstrate the problems involved, suppose we treat by these various methods the small group of Ra-226 results actually obtained in the intercomparison exercise discussed above. Suppose we further make the entirely reasonable assumption that some of the laboratories are not as competent in their analytical capabilities as the others and sometimes, as can be seen, fail completely for whatever reason. Using the root-mean-square equation, the mean and experimental standard deviation for the 10 results are  $2.91 \pm 2.48$  pCi/g. This method is fine for showing the dispersion of the results obtained from the contributing laboratories, but is completely unacceptable for determining a best value for the sample. It democratically gives equal weight to each laboratory regardless of the precision with which the

measurement was made or the obvious fact that some of the laboratories are out of control. Similarly, to attempt to discard some results as outliers from such a small population of agreeable results is an exercise in futility.

Using the reciprocal of the variance as a weighting factor as recommended by others, the weighted mean and standard deviation for the same 10 results are  $0.204 \pm 0.009$  pCi/g, a completely unacceptable result. Over 90% of the total weight is given to the single result most clearly known to be incorrect, i.e.,  $0.00 \pm 0.01$ , simply because the absolute value of its standard deviation is the smallest. On the other hand, using the reciprocal of the square of the relative standard deviation as the weighting factor as recommended above, the weighted mean and standard deviation are  $2.66 \pm 0.04$  pCi/g. Of the total weight, 14.6, 47.6 and 35.8% were contributed by the results  $5.5 \pm 0.2$ ,  $2.48 \pm 0.05$ , and  $1.72 \pm 0.04$ , respectively, as they should have been, based on the relative standard deviations involved. Unfortunately, statistics can not help us decide which are the more accurate results. As mentioned above, this must be demonstrated before the intercomparison is carried out.

There is a corollary that must be derived from this exercise that is very far-reaching in its implications. If the democratic process is to be observed by giving equal weight to results from participating laboratories in round-robin intercomparisons in an

attempt to improve the accuracy of the results, all measurements must be made with approximately the same precision. This will be difficult to accomplish but the effort must be made. Otherwise, a 20-laboratory collaborative effort might result in a conclusion based on the results from the single laboratory producing the most precise (but not necessarily the most accurate!) measurement.

## REFERENCES

1. C. W. Sill, "Critique of Current Practice in the Determination of Actinides," Proceedings of a Workshop on Measurement and Interpretation of Actinide Accumulation by Man, University of Utah, Salt Lake City, Utah, 1979, in press.
2. R. G. Beverley, Union Carbide Corporation, Grand Junction, Colorado, private communication, 1979.
3. G. G. Mastinu, "Quality Control Procedures Applied to Environmental Radioassay in Italy," Health Phys., 36, 47-52 (1979).
4. R. Dybczynski, A. Tugsavul, and O. Suschny, "Problems of Accuracy and Precision in the Determination of Trace Elements in Water as Shown by Recent International Atomic Energy Agency Intercomparison Tests," Analyst, 103, 734 (1978).
5. R. P. Bernabee, D. R. Percival, and D. B. Martin, "Fractionation of Radionuclides in Liquid Samples from Nuclear Power Facilities," Health Phys., 39, 59-67 (1980).
6. C. W. Sill, K. W. Puphal, and F. D. Hindman, "Simultaneous Determination of Alpha Emitting Nuclides of Radium Through Californium in Soil," Anal. Chem., 46, 1725-1737 (1974).

7. C. W. Sill and R. L. Williams, "Preparation of Actinides for Alpha Spectrometry Without Electrodeposition," in publication.