

**APPENDIX C**

**STATISTICAL METHODS FOR INTERPRETING  
BIOASSAY DATA**

## APPENDIX C

### STATISTICAL METHODS FOR INTERPRETING BIOASSAY DATA

Inevitably statistical methods must be applied to bioassay data to provide meaning to the data and to use the data in models. The more important statistical methods used by the Hanford Internal Dosimetry Program are described below.

#### DECISION LEVEL AND MINIMUM DETECTABLE AMOUNT

Three fundamental questions in dealing with bioassay data are 1) when does a sample result indicate the presence of something (the analyte)? i.e., when is the analyte detected? 2) what is the overall capability of the bioassay method for continual assurance of detection of the analyte?, and 3) when does a sample result indicate the presence of analyte above the natural background? The third question is important mostly for uranium in excreta and is discussed in the section on uranium. Concerning the first two questions, the Hanford Internal Dosimetry Program follows the concepts of decision level (also called critical level) and minimum detectable amount (also called detection level or lower level of detection) as described by Currie (1968, 1984), Brodsky (1986), the American National Standards Institute<sup>(a)</sup>, and many others.

The decision level,  $L_c$ , is the parameter that is used to answer question 1) above. The  $L_c$  is dependent on the probability of obtaining false positive results (or probability of a type I error) that one is willing to accept. For indirect bioassay samples, a 5% probability of false positives is used to calculate the  $L_c$ . The  $L_c$  is calculated from the results of analyses of blank

---

(a) American National Standards Institute (ANSI). 1987. Performance Criteria for Radiobioassay. Draft ANSI Standard N13.30, New York, New York.

samples. (b) Because the actual  $L_c$  value for a given radionuclide analysis will vary slightly over time, a single value is then chosen to define "positive" samples. This value, called positive level, is set to be slightly greater than the temporal fluctuations of the short-term  $L_c$  values and is a number that is convenient to use by all concerned parties. Any sample result that is equal to or greater than the positive level is considered positive, i.e., the analyte has been detected. It is inherent in this method that when dealing with large numbers of samples many samples containing no activity will be declared positive, those being somewhat less than 5% of the total unless a subtle bias exists in the population of samples.

The minimum detectable amount (or activity), MDA, is the parameter that answers question 2) above. The MDA is a function of the probabilities of both false positive and false negative (type II error) results. For indirect bioassay results, the probability of each kind of error is set at 5%. The MDA is determined annually from analysis of blank samples. Annual MDAs are compared with values set by contract with the bioassay laboratory. (The MDAs must be less than contractual detection levels or corrective action is immediately undertaken.) It is the contractual detection levels that are referenced throughout this document because only the contractual detection levels are enforceable and are generally applicable over long periods of time. At any time though, actual MDAs are usually somewhat lower than the contractual detection levels quoted in this document.

For in vivo measurements, because each person provides a different background spectrum resulting from different amounts of natural radionuclides, detectability is determined using information from both blank subjects (with normal activities of  $^{40}\text{K}$ ) and from the person being counted. Background counts in the region of interest in the photon energy spectrum are determined by measuring the counts in nearby channels. From a group of blank subjects, the ratio between the counts in the region of interest and the counts in nearby channels is determined. This ratio is then applied to the counts in the nearby channels of the spectrum from the person being counted to determine

---

(b) Blank samples are those that are equivalent to an actual sample but are known not to contain the analyte (uranium excepted).

the background in the region of interest. The standard deviation of the counts in the nearby channels is also calculated. If the total counts in the region of interest are greater than the background counts plus K times the standard deviation, then a photopeak in the region of interest is considered "detected." K is 3 for spectra obtained using germanium detectors and 4.6 for spectra obtained using NaI detectors. Except at very low energies, the ratio of nearby-channel counts to background-region-of-interest counts is considered to be 1 for spectra obtained using germanium detectors. Consideration is given to the presence or absence of accompanying photopeaks and possible interference from other radionuclides before a radionuclide is considered "detected." Application of complete Currie-type statistics to in vivo counting is under development.

#### NORMALIZATION OF INDIRECT BIOASSAY DATA

Indirect bioassay data may be normalized differently according to the type and processing category of the sample. Table C.1 describes the types of samples and Table C.2 describes the processing categories.

TABLE C.1. Types of Indirect Bioassay Samples

<u>Media</u>	<u>Description</u>
Urine	Simulated 24-hour collection. Total excretion between just before sleeping and just after rising on 2 consecutive days. Not collected at work (NCRP 1987).
Urine	Total 24-hour collection, including collection at work if necessary.
Urine	Simulated 12-hour collection. Total excretion between just before sleeping and just after rising; one time, not at work.
Urine	Spot sample. A single voiding.
Feces	Single voiding.

TABLE C.2. Processing Categories

<u>Name</u>	<u>Description</u>
Routine	Longest turnaround time, most sensitive analysis techniques. Used for most routine samples and samples for which turnaround time is not critical.
Priority	Shorter turnaround time but same sensitivity as routine processing. Generally used for sampling terminating workers and for following up on potential intakes.
Expedite	Three-day turnaround time, intermediate sensitivity. Used for follow-up on a potential intake, usually shortly after the intake.
Emergency	Turnaround in a few hours, least sensitivity. Used for rapid indication of the seriousness of an intake.

Generally, the urine data are normalized to total 24-hour excretion. Provided the sample is collected properly, a total or simulated 24-hour urine sample result is used as is; no further normalization is done. A proper 12-hour sample result is normalized by doubling the result.

If it is suspected that a sample has not been provided according to instructions, several approaches are taken. A routine-category sample that is supposed to contain 24-hour excretion is not analyzed if the volume is less than 500 mL because the volume is too small to represent a true 24-hour collection. The worker is asked to provide another sample. The minimum volume for other routine-category samples and for all priority-category samples is 20 mL. When a sample is above the cutoff for analysis but is believed not to represent 24-hour (or 12-hour) excretion, the result may be 1) normalized to 24-hour based on information from the provider, 2) ignored, 3) normalized by volume to 24-hour excretion, or 4) normalized by creatinine to 24-hour excretion. To normalize by volume, 1400 mL for males and 1000 mL for females (from Reference Man [ICRP 1974]) should be used for 24-hour excretion unless the person-specific daily excretion rate is known. Normalization by creatinine should be used only if the person-specific daily creatinine excretion is established.

The one exception to the above discussion concerns analysis for tritium in urine. Spot urine samples are collected, and the data are reported and used directly in units of activity per volume (see Section 2.0).

Concerning fecal samples, total fecal excretion over a specified period of time (e.g., 3 or 5 days) is obtained. If it appears that the person has not followed instructions, attempts are made to determine what portion of the time is represented by the samples collected. As a last resort, a single voiding is assumed to represent 24-hour excretion. (See also Appendix E.)

#### COMPARISON OF RETENTION OR EXCRETION FUNCTIONS TO DATA

Generally, following an acute intake, more than one datum of a similar nature (e.g., urine results) is obtained, and the appropriate retention or excretion function is fit to the data. Two options for fitting the function to the data are used ([listed in order of preference]: 1) weighted least-squares fit, 2) unweighted least-squares fit), and other methods are being investigated.

The weighted least-squares fit should be used when two results of the measurement process are known--the result itself (whether zero, negative, or whatever),  $x_i$ , and its variance,  $\sigma_i^2$ --and when the variances are all determined in the same manner. The weighting factor is the inverse of the sum of the variances. The intake is given by

$$I = \left[ \sum_{i=1}^n \frac{x_i r_i}{\sigma_i^2} \right] / \left[ \sum_{i=1}^n \frac{r_i^2}{\sigma_i^2} \right] \quad (C.1)$$

where  $r_i$  is the value of the fractional retention or excretion function at the same time after intake as the sample result  $x_i$  (Bevington 1969). Use of the weighted least-squares fit avoids having the calculation of intake or uptake dominated by a few large data points that may have poor precision, such as a hastily analyzed urine sample collected shortly after an intake.

If the variances are unknown, are known to be equal, or were determined differently (such as counting uncertainty versus total propagated

uncertainty), then the unweighted least-squares fit should be used. The unweighted least-squares fit is represented by Equation (C.1) when all variances are set equal to one.

Data that are listed only as "less than" some value are difficult to use in a mathematical fitting technique. The Hanford Internal Dosimetry Program arbitrarily sets the value for the measurement as one-half of the less-than value for use in least-squares fitting techniques. This does not work well if too many of the data are less-than values. If there are many less-than values and a few well-known data, then the evaluator may need to use only the well-known data in the least-squares fitting technique, making sure that the best fit does not seem unreasonable with regard to the many less-than data.

In all cases, outliers, or data that are not relevant to the equation being fit, should not be included in a fitting technique. Examples would include urine data influenced by diethylene triamine penta acetate (DTPA) therapy or a datum with a very high less-than result. The assessment should document which data are being ignored and why.

#### TREATMENT OF RECOUNTED DATA BEFORE USING IT WITH ONCE-COUNTED DATA

Results from samples that have been recounted should not be used directly with results from once-counted samples in analysis programs such as PCPLOT. The best estimate of the mean value of the recounted sample and the best estimate of the uncertainty of the mean value need to be determined first so that each sample has only one value. The mean value ( $\bar{x}$ ) should be determined by the formula for a weighted average:

$$\bar{x} = \frac{\sum_{i=1}^n \frac{X_i}{S_i^2}}{\sum_{i=1}^n \frac{1}{S_i^2}} \quad (n \text{ is usually } 2 \text{ or } 3) \quad (C.2)$$

where  $X_i$  are the results from the recounts and  $S_i$  are the total propagated uncertainties.

The best estimate of the uncertainty of the mean value should be determined by the formula for a weighted uncertainty:

$$\bar{S} = \left[ 1 + \left( \sum_{i=1}^n \left( \frac{1}{S_i^2} \right) \right) \right]^{1/2} \quad (C.3)$$

This approach provides consistency in the way recounted sample data are used in dose assessments, and prevents recounted data from acquiring unwarranted weight relative to once-counted data for curve-fitting purposes.

#### REFERENCES

Bevington, P. R. 1969. Data Reduction and Error Analysis for the Physical Sciences, pp. 69-73. McGraw-Hill Book Company, New York, New York.

Brodsky, A. 1986. Accuracy and Detection Limits for Bioassay Measurements in Radiation Protection. NUREG-1156, U.S. Nuclear Regulatory Commission, Washington D.C.

Currie, L. A. 1968. "Limits for Qualitative Detection and Quantitative Determination." Analytical Chemistry 40(3):586-593.

Currie, L. A. 1984. Lower Limit of Detection: Definition and Elaboration of a Proposed Position for Radiological Effluent and Environmental Measurements. NUREG/CR-4007, U.S. Nuclear Regulatory Commission, Washington D.C.

National Council on Radiation Protection and Measurements (NCRP). 1987. Use of Bioassay Procedures for Assessment of Internal Radionuclide Deposition. NCRP Report No. 87, Bethesda, Maryland.