B Level 1 Heading

THIS PAGE CONTAINS A LEVEL 1 HEADING WHICH ACTIVATES THE APPENDIX LETTER IN THE PAGE NUMBER. THE HEADING 1 STYLE MUST BE MODIFIED IN ORDER TO CHANGE THE APPENDIX LETTER.

PLEASE RECYCLE THIS PAGE.

Appendix B

Statistical Methods for Internal Dosimetry

Appendix B

Statistical Methods for Internal Dosimetry

The Hanford Internal Dosimetry Program (HIDP) uses statistical methods to interpret bioassay data. Some of the principal statistical methods used by the HIDP are described in this appendix, including those used to 1) determine when a sample result indicates the presence of something (i.e., when the analyte is detected), 2) describe the overall capability of the bioassay method for continual assurance of detection of the analyte, 3) normalize data, and 4) fit data to retention or excretion functions to calculate intake. An additional issue is the determination that a detected analyte is significantly different from the normal presence of that analyte in a sample due to natural background. This latter item is particularly important for uranium in urine and is discussed in that context in Chapter 7.0.

The HIDP follows the concepts of critical level for decision, L_c, (also called the decision level, DL) and minimum detectable amount (or activity) MDA (also called detection level, L_d, or lower limit of detection, LLD in various publications), as described by Currie (1968; 1984), Brodsky (1986), the Health Physics Society (HPS 1996), and many others.

B.1 Decision Level

The L_c is the parameter that is used to indicate that an analyte has been detected. The L_c is dependent on the probability of obtaining false positive results (i.e., the probability of a type I error) that one is willing to accept. Decision levels for in vivo measurements are determined as described in the *In Vivo Monitoring Program Manual*.[(a)] Decision levels for excreta samples are described below.

Until April 2000, the decision levels by which excreta samples containing elevated quantities of radioactive material were identified were all specified in the contract with the analytical laboratory as absolute activity values. The contract specified an upper bound for

(a) Pacific Northwest National Laboratory (PNNL). *In Vivo Monitoring Project Manual.* PNNL-MA-574, Richland, Washington. (Internal manual.)

the MDA for each analysis—called the contract limit, CL—and established the decision level at one-half the CL. For most alpha spectrometry analyses the CL was 0.02 dpm per sample. The decision level was applied to the activity result, rather than the count data. The detection criterion was therefore insensitive to sample specific variables such as chemical yield and detector efficiency.

In order to select a better procedure, an investigation of the American National Standard HPS N13.30 (1996) and other proposed decision level equations was initiated. The investigation concluded that the N13.30 equation significantly underestimates the number of false positive results (MacLellan 2000). The maximum number of false positive results peaks at about one background count during the counting period, but remains significant up to an expected 100 counts. The N13.30 equation answers the question, "How large a net count will be expected, less than 'alpha' percent of the time, for a given background count rate if there is no activity in the sample?" An equation proposed by Altshuler and Pasternack (1963) was found to be far superior:

$$\Delta_1 = k_\alpha \sqrt{2n_b} \left(\frac{k_\alpha}{\sqrt{8n_b}} + \sqrt{1 + \frac{k_\alpha^2}{8n_b}} \right) \qquad \text{(B.1)}$$

where $)_1$ is the decision level, k_α is the false positive error term, and n_b is the background count. That equation answers the question "How much activity can be in a sample, and the confidence interval for the net count still include zero?" The decision level obtained from the Altshuler and Pasternack equation remains unbiased down to an expected three counts during the counting period.

For alpha spectrometry analyses, the concept of the Altshuler and Pasternack equation was adopted for the Hanford bioassay program, but an additional simplification was incorporated. Although the form of the equation investigated previously calculates the decision level based only on the number of expected background counts, the original form of the equation used in the derivation equates the decision level with a multiple of the standard deviations estimate of a net count value $()_1$). That is, the decision level is derived from the following equation:

$$\Delta_1 = k_\alpha \sqrt{n_g + n_b} = k_\alpha \sqrt{\Delta_1 + 2n_b} \qquad \text{(B.2)}$$

where $(n_g + n_b)$ is used as the estimate of the net count variance, n_g is the gross count, and n_b is the background count. Equation B.1 was derived by solving Equation B.2 for $)_1$. Hanford denotes the decision level by L_c, rather than the $)_1$ used in Equations B.1 and B.2.

Equations B.1 and B.2 are based on counts, however, the Hanford bioassay contract requires an estimate of the total propagated uncertainty (TPU) for each result reported. A sample specific decision level has been implemented based on the TPU. Rather than using Equation B.1, Hanford alpha spectrometry decision levels are set from Equation B.2, substituting the TPU for the radical. A k_∇ value of 2 was chosen in order to maintain the average decision level near historic levels. The decision level, L_c, then takes the following simple form:

$$L_c = 2 \times \text{TPU} \tag{B.3}$$

Any sample result that is equal to or greater than the L_c is considered positive, i.e., the analyte has been detected. It is inherent in this method that when dealing with large numbers of samples some samples containing no activity will be declared positive. Using the above criteria, about two percent of the results are expected to be false positives assuming a normal distribution of the net count.

B.2 Minimum Detectable Amount

The minimum detectable amount (or activity), MDA, provides a statement of the overall capability of the bioassay method to provide continual assurance of analyte detection. The MDA is a function of the probabilities of both false positive and false negative (type II error) results. For excreta bioassay results, the probability of each kind of error is set at 5%. The MDA is determined annually from analysis of blank samples, using the methods of HPS N13.30 (1996). 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 required.) 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.

Determinations of in vivo measurement MDAs are described in the *In Vivo Monitoring Program Manual*.[(a)]

B.3 Normalization of Excreta Bioassay Data

Excreta bioassay data may be normalized differently according to the type of the sample. Generally, urine data are normalized to total 24-hour excretion based on the sampling protocol. Total urine sampling for 24 hours does not require normalization, however, that protocol is generally not convenient for workers. Thus, an approximate 24-hour sample is generally used. The approximate 24-hour sample protocol (historically referred to as a "simulated 24-hour collection") involves collecting all urine voided between one-half to one hour before retiring at night through the first voiding after getting up in the morning, this done for two consecutive nights. Provided the sample is collected properly, a total or approximate 24-hour urine sample result is used as is; no further normalization is done. An overnight sample is considered to represent approximately 12 hours of urine collection, and is normalized by doubling the result. Alternate normalization methods are described in Chapter 2.0, and may be used if they are more appropriate for the actual data.

B.4 Treatment of Recounted Data Before Using It with Once-Counted Samples

Results from samples that have been recounted should not be used directly with results from once-counted samples in analysis programs such as CINDY. 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 (x_{avg}) should be determined by the following formula for a weighted average:

$$x_{avg} = \frac{\sum_{i=1}^{n} \frac{x_i}{s_i^2}}{\sum_{i=1}^{n} \frac{1}{s_i^2}} \quad \text{(n is usually 2 or 3)} \qquad \text{(B.4)}$$

(a) Pacific Northwest National Laboratory (PNNL). *In Vivo Monitoring Project Manual.* PNNL-MA-574, Richland, Washington. (Internal manual.)

where x_i is each measurement result and s_i is the total propagated uncertainty associated with the measurement.

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

$$s_{avg} = \sqrt{\frac{1}{\sum_{i=1}^{n} \frac{1}{s_i^2}}} \tag{B.5}$$

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.

B.5 Curve-Fitting Techniques

When multiple data of a similar nature (e.g., urine results) are obtained following an intake, some kind of curve fit is performed to fit an appropriate retention or excretion function to the data. The CINDY computer code (Strenge et al. 1992) is most commonly used for the fitting of data, and the fitting algorithms available in that code are briefly discussed in Appendix D. A recent study has shown that the "average of the slopes" fit method in CINDY is preferred if the predominant variance comes from the biology and not sample analysis (Skrable 2000). Alternatively, simple data fits may be accomplished using one of the following techniques.

The weighted least-squares fit is appropriate when two results of the measurement process are known—the result itself (whether zero, negative, or positive), x_i, and its variance, σ^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

$$I = \frac{\sum_{i=1}^{n} \frac{x_i \, r_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{r_i^2}{\sigma_i^2}} \tag{B.6}$$

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 (or uniform weighting) least-squares fit is appropriate for use. The unweighted least-squares fit is represented by Equation B.6 when all variances are set equal to one, or as shown below:

$$I = \frac{\sum_{i=1}^{n} x_i \; r_i}{\sum_{i=1}^{n} r_i^2} \qquad \text{(B.7)}$$

Data that are listed only as "less than" some value are difficult to use in a mathematical fitting technique. At times, the HIDP has arbitrarily set 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 dosimetrist 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.

An "eye-ball" fit consists of plotting a line through a data set and graphically extracting the required intercepts and slope. This approach involves subjective judgment by the dosimetrist and is not an objective statistical fit. An "eye-ball" fit may be necessary if too many data are "less-than" values or if the analytical sensitivity varied greatly from datum to datum. Caution must be used in exercising eye-ball fits, because the quality of data obtained will vary depending on the type of axis (linear, logarithmic).

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

acetic acid (DTPA) therapy or a datum with a very high less-than result. The assessment should document which data are being ignored and why.

B.6 References

Altshuler, B., and B. Pasternack. 1963. "Statistical measures of the lower limit of detection of a radioactivity counter." *Health Phys*. 9:293-298.

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.

Health Physics Society (HPS). 1996. *Performance Criteria for Radiobioassay*. HPS N13.30-1996. An American National Standard approved by American National Standards Institute, Health Physics Society, McLean, Virginia.

MacLellan, J. A. 2000. *Hanford's Decision Level for Alpha Spectrometry Bioassay Analyses Based on the Sample Specific Total Propagated Uncertainty*. PNNL-SA-33598, Pacific Northwest National Laboratory, Richland, Washington.

Skrable, K. 2000. "Variance models for estimating intakes from bioassay data." HPS Newsletter, XXVIII (4):18-20. Health Physics Society, MacLean, Virginia.

Strenge, D. L., R. A. Kennedy, M. J. Sula, and J. R. Johnson. 1992. *Code for Internal Dosimetry (CINDY Version 1.2), Part 1: Conceptual Representation*. PNL-7493 Pt. 1, Rev. 1., Pacific Northwest National Laboratory, Richland, Washington.