7. DATA ANALYSIS AND STATISTICAL TREATMENT

7.1 INTRODUCTION

Four goals for analysis and statistical treatment of data are identified in Chap. 7 of the *Regulatory Guide*. These goals address estimation of radionuclide concentrations and measures of accuracy and precision, comparisons to previous values and regulatory or dose limits, comparison of on-site concentrations to reference location concentrations, and evaluations of the comparisons. These goals are addressed for both the radiological and nonradiological monitoring programs through a general discussion of the DQO process, data analysis, and by addressing the *should** elements in the *Regulatory Guide* section.

7.2 DATA QUALITY OBJECTIVES

The principal means of ensuring that data are of sufficient quality and adequate quantity to achieve a monitoring or surveillance objective is to establish DQOs for the program. These are the specifications for the sampling, laboratory analysis, and data evaluation that support decisions or comparisons to criteria. Considerations in the DQO development process include the types of decisions that will be supported by the data, how the data will be used, requirements for accuracy and precision of methods, acceptable levels of uncertainty, comparability to previously collected data, representativeness, and repeatability. The process is a collaborative effort that requires the involvement of all the disciplines that have responsibility in the path from sampling design through the final decision or estimate.

7.3 DATA ANALYSIS

The approach to data analysis at each of the three facilities on ORR has evolved in response to effluent monitoring and environmental surveillance requirements, reporting schedules, and available resources. Effluent monitoring data are assessed against regulatory requirements and evaluated for trends on a schedule that is consistent with reporting requirements. These requirements are identified in the implementing regulations of the laws and site-specific permits.

All effluent monitoring and environmental surveillance data are summarized in ASER. The general purpose of this information is to chronicle the impacts of the operations at the facilities and not to identify changes in operations or process upsets when they occur. Environmental monitoring programs are intended to assess impacts much more frequently than to identify and report upset conditions. In the event that an effluent parameter shows elevated values, the environmental surveillance data may be assessed in an attempt to corroborate the effluent results.

7.4 *REGULATORY GUIDE* PERFORMANCE CRITERIA

a. The statistical techniques used to support the concentration estimates, to determine their corresponding measures of reliability, and to compare radionuclide data between sampling and/or measurement points and times *should be designed with consideration of the characteristics of effluent and environmental data.**

A majority of the data presentations used to meet regulatory reporting requirements and summarize surveillance data consists of averages, maxima, minima, standard deviations, and/or standard error. The vast majority of information that has been reported and/or published in the past has been summarized using these standard techniques. The intent of these summaries is to show "typical" value and range of data. It is worthwhile to continue to provide data in this manner because the mean and range meet the intent of the summaries and continued consistency in reporting supports comparisons of information across multiple years.

Statistical techniques such as analysis of variance, correlation analysis, or regression may warrant an evaluation of characteristics of skewness, kurtosis, and additivity against the assumptions and robustness of the methods. Whenever small amounts of data are available, statistical judgements should not be followed blindly. Whenever there is conflict between an intuitive and statistically-based conclusion, the basis for the conclusions should be reviewed to ensure a relevant factor has not been overlooked in the reasoning process or determine if the statistical model is faulty. Face-value comparisons using trends, percent differences between duplicates, and professional judgement are often the preferred methods.

Effluent monitoring data are typically compared to permit limits or regulatory standards in a pass/fail manner. Rejecting an effluent datum as an outlier is usually not an option under the permit or regulation. In some instances regulatory agencies have provided guidance for statistical analysis of data from environmental media [i.e., EPA "Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities," (EPA/530-SW-89-026, 1989)]. Where such guidance exists, it is applied as appropriate for the media and regulatory program for which it is directed.

b. Documented and approved sampling, sample-handling, analysis, and data management techniques *should be used to reduce the variability of results.**

The techniques for sampling, sample handling, chain of custody, instrument maintenance and calibration, laboratory analysis, data management, information analysis, and reporting are documented in the SOPs of the organizations that perform the work. All compliance environmental monitoring is conducted in accordance with the specific regulatory authorities that mandate the activity. Sampling, chain of custody, and laboratory analyses are conducted in accordance with requirements and guidance from EPA and DOE, as available.

Laboratory analyses are conducted in accordance with controlled procedures that are resident in the analytical laboratory. Information regarding these procedures is located in Sect. 5 of this EMP.

c. The level of confidence in the data due to the radiological analyses *should be estimated by analyzing blanks and spiked pseudosamples and by comparing the resulting concentration estimates to the known concentrations in those samples.**

The Y-12 laboratory participates in the EPA Intercomparison Radionuclide Control Program administered by the EPA Environmental Monitoring System Laboratory at Las Vegas and the DOE Environmental Measurements Laboratory Radionuclide Quality Assessment Program. These two programs provide pseudosamples of a variety of media for analysis. Blanks are routinely run as part of the laboratory QC program.

d. The precision of radionuclide analytical results *should be reported as a range, a variance, a standard deviation, a standard error, and/or a confidence interval.**

All radioanalytical data are reported as the value ±2 times the estimated standard deviation of the value, based on counting statistics.

e. Data *should be examined and entered into the database promptly after analysis.**

Analytical data are managed in the analytical laboratories electronically on information management systems. These data are electronically transferred to the information management organizations that support each site.

Field measurements are entered into the database in accordance with the SOPs of the responsible organization.

f. Outliers *should be excluded from the data only after investigation confirms that an error has been made in the sample collection, preparation, measurement, or data analysis process. As each data point is collected, it *should** be compared to previous data, because such comparison can help identify unusual measurements that require investigation or further statistical evaluation.**

The analysis of data for outliers is primarily relevant to surveillance data; effluent monitoring data are typically compared to permit limits or regulatory standards in a pass/fail manner. Rejecting an effluent datum as an outlier is usually not an option under the permit or regulation.

The surveillance data verification process includes comparing new data to the 5th percentile and 95th percentile of the 2-year range for each analyte. This approach avoids using an outlier or extreme value as a data qualifying limit. Range limits may be established across an entire program, subsets of a program, or for each monitoring point within the program.

Values that exceed the range are automatically verified with the analytical laboratory. If the laboratory cannot identify a measurement or data management error, then the field records associated with the sample are checked for any annotations regarding unusual conditions that might affect the results.

Outliers are not excluded from information analysis unless there is a declarative reason to do so. Declarative reasons include identifying the source of the error that generated the unusual result or a value that is not possible, such as a pH of 17.3. This approach is taken because of the stochastic nature of effluent and environmental data; sometimes the unusual result is an important indicator of a change in effluent or environmental conditions. A suspicious value is noted in the ASER text that accompanies the data summary. If an outlier is identified and a valid measurement is available from another source, the outlier may be replaced. An explanation of why the substitute is acceptable will be given in ASER.

g. As they apply to data analysis and statistical treatment activities, the general QA program provisions of Chap. 10 of this guide *should be followed.**

The activities discussed in this section are conducted by organizations that have QAPs that are consistent with the requirements for environmental monitoring. Additional information is provided in Sect. 9 of this EMP.