CONTAMINATED SITES STATISTICAL APPLICATIONS GUIDANCE DOCUMENT NO. 12-11

STATISTICAL QA/QC

A guide for project managers, reviewers, data analysts and interpreters on statistical quality assurance and quality control

This guidance document is one of a series that outlines important basic statistical concepts and procedures that are useful in contaminated sites studies. BC Environment recommends that these suggestions be followed where applicable, but is open to other techniques provided that these alternatives are technically sound. Before a different methodology is adopted it should be discussed with BC Environment.

April 2001

THE GENERAL IDEA

A statistical study is only as reliable as the data on which it is based; if the fundamental data are called into question, the entire study and its conclusions are also called into question. It is important, therefore, to be able to document how reliable the data are. Issues related to the reliability of data are often grouped under the general heading of "quality assurance and quality control" (QA/QC), a description that captures the idea that data quality can not only be documented but can also be controlled through appropriate practices and procedures.

Even with the most stringent and costly controls, data will never be perfect: errors are inevitable as samples are collected, prepared and analyzed. One goal of QA/QC is to quantify these errors so that subsequent statistical analysis and interpretation can take them into account. A second goal is to monitor the errors so that spurious or biased data can be recognized and, if possible, corrected. A third goal is to provide information that can be used to improve sampling practices and analytical procedures so that the impact of errors can be minimized.

This guidance document begins with a discussion of two concepts: accuracy and precision. It then presents statistical tools that can be used to study the accuracy and precision of existing data, and also presents ideas on practices and procedures that allow accuracy and precision to be monitored as the data are being collected. It closes with a brief discussion of some aspects of QA/QC that are often overlooked: the reliability of location information, the reliability of qualitative information and the reliability of computerized data bases.

ACCURACY AND PRECISION

Statistical QA/QC involves two separate but related concepts: accuracy and precision. Figure 1 captures the difference between these two concepts. A sample is accurate if repeated attempts are centered about the target value; it is precise if repeated attempts are all close to one another.

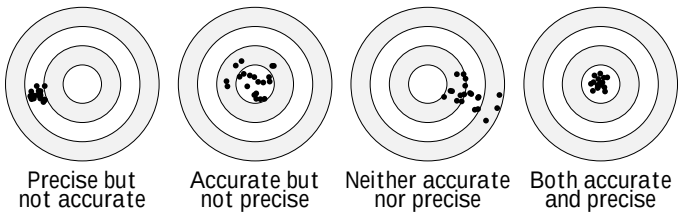

Figure 1 Examples of accuracy and precision.

For the specific case of analytical values, where repeated measurements of the same sample are possible (though somewhat expensive), we can imagine an experiment in which we reanalyze the same material 100 times. Figure 2 shows how accuracy and precision might manifest themselves on a histogram of the repeat analyses; in this example, the target we are aiming for is the true PCB concentration of 10 ug/g.

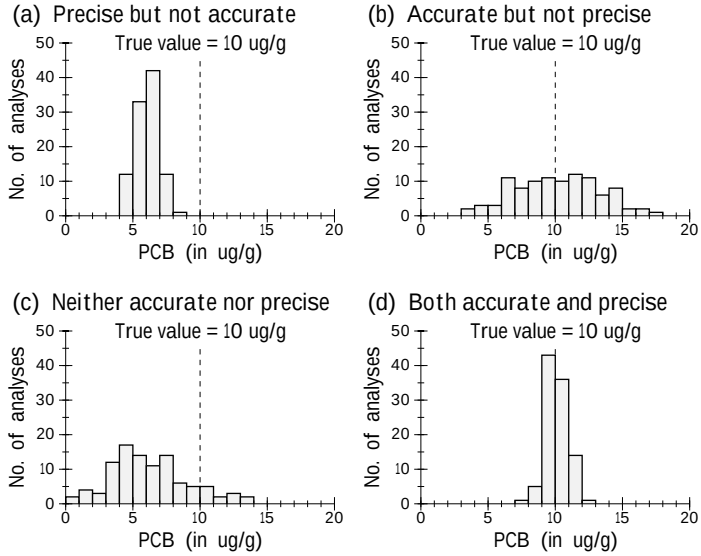

Figure 2 Histograms of 100 repeat analyses.

If the analytical procedure is inaccurate then the average of repeat analyses will be different from the true value; the difference between the average of repeat analyses and the true value is often called the bias. A histogram of repeat analyses from an inaccurate procedure will not be centered about the true value, as in Figures 2a) and 2c). If the analytical procedure is imprecise then repeat analyses will not be close to one another. As precision improves, the spread of the histogram of repeat analyses will decrease; Figures 2a) and 2c) both show repeat analyses that are inaccurate, but those in Figure 2a) are more precise because they show less scatter. Precise analyses are often referred to as "repeatable" because repeated analyses all come close to the same value. As Figure 2a) shows, precise or repeatable analyses are not necessarily accurate and may simply be coming close to the same wrong value.

Though the example in Figure 2 is built on repeat analyses of the same material, it should be noted that sampling errors are not solely due to the analytical procedure. The earlier steps of sample collection and sample preparation often contribute more to the total error than the analytical procedure used in the laboratory. Statistical QA/QC should attempt to document and control the accuracy and precision of each step in the sampling procedure, from the initial extraction of the material from the ground to the final analytical value produced by the lab.

Inaccuracy or bias in a sampling procedure is due to systematic errors that cause the sample values to be generally too high or too low. Imprecision, on the other hand, is a result of random errors that do not have any systematic bias but that cause the sample value to be different from the true value.

MISCLASSIFICATION

Though statistical QA/QC traditionally deals with accuracy, as reflected in the mean value of repeat analyses, and precision, as reflected in the variance of repeat analyses, these may not be the most critical statistical characteristics for remediation planning. For much of the data collected from contaminated sites, their purpose is to determine whether material is contaminated or not and our primary concern should be whether the sample values are above or below some critical threshold. An inaccurate or imprecise sampling procedure may have little consequence if it does not cause contaminated material to be misclassified as uncontaminated or vice versa.

Though improvements in precision and accuracy usually go hand-in-hand with improvements in classification, this is not necessarily the case. Statistical QA/QC for contaminated site studies should not focus exclusively on accuracy and precision but should also address the issue of misclassification.

MONITORING AND CHECKING DATA QUALITY

Control charts for reference standards

An ideal approach to studying the reliability of an analytical procedure is to reanalyze a prepared standard whose true value is known. Such reference material can be specifically prepared for a particular site; for many contaminants, reference material is also available commercially. The advantage of site-specific reference material is that its chemical and physical composition is representative of the particular site; if an analytical procedure is known to be sensitive to factors such as moisture or clay content, then site-specific reference material will provide the best opportunity for calibrating the analytical procedure. The advantage of commercial standards is that their true value has been well established; they have either gone through a battery of different analytical procedures by different laboratories or have been carefully prepared by spiking uncontaminated material with known concentrations of the contaminant.

At regular intervals during the course of a project, the reference material can be included for analysis along with other samples. The resulting repeated analyses of the reference material can then be plotted on a control chart that shows how the analytical values of the reference material fluctuate with time.

Figure 3 shows a control chart for reference material that was prepared for a site contaminated with mercury. With several hundred samples being collected and analyzed at an on-site laboratory over the space of a few weeks, it was decided that the reference material should be checked daily, so the control chart in Figure 3 shows one analytical value per day (except Sundays and some Saturdays). The dashed line on Figure 3 shows the accepted true value for the reference material, which was prepared commercially and was designed to have a true value of 20 ug/g.

In addition to showing the accepted reference value, a control chart should also show the range of acceptable fluctuations around this reference value. The limits for acceptable analytical values can either be established through an initial batch of repeat analyses or can be dictated by remedial design objectives. For the control chart shown in Figure 3, the minimum and maximum acceptable values are shown as dotted lines and are based on design objectives that require the analytical values to be within $\pm 10\%$ of the true value.

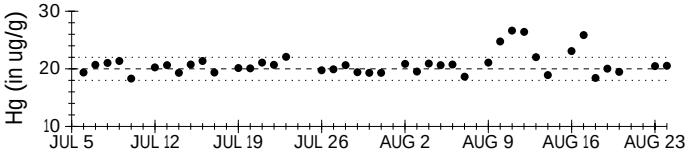

Figure 3 Example of a control chart.

The control chart in Figure 3 shows that the analytical procedure was generally acceptable in terms of its accuracy and precision. For a short period of time towards the end of the sampling exercise, the analyses became somewhat biased and more erratic. In this particular case, it took several days to determine that the cause of these unacceptable errors was operator error but, once identified, these errors were easily corrected. All samples that were initially analyzed during the troublesome period were reanalyzed to provide more reliable analytical values for remediation planning.

Though control charts are excellent for monitoring data quality, they generally focus on the analytical errors that accumulate after a sample has been collected and prepared. The errors that occur in the collection of the original sample material and the preparation of the subsample that is finally analyzed are often much greater than those that occur in the actual analysis of the prepared material. Even though control charts may show acceptable accuracy and precision, a thorough QA/QC program should also investigate errors that occur before the final prepared subsample is delivered to the analytical device.

Scatterplots and summary statistics for paired analyses

Another method for checking the quality of analytical data is to reanalyze several samples and to do a statistical study of the paired analyses using scatterplots and a few summary statistics. The statistical differences between the two sets of analyses will reflect the cumulative effect of all the differences in the way that the two sets of samples were collected, prepared and analyzed. A few examples of some different types of sample pairs will illustrate some of the different combinations of errors that such a study might address:

- The paired values can be one lab's reanalyses of the same prepared material, in which case the statistical comparison will reflect intra-laboratory analytical errors.
- The paired values can be reanalyses performed by different laboratories of the same prepared material, in which case the statistical comparison will reflect inter-laboratory analytical errors between labs.
- The paired values can be reanalyses performed by the same laboratory from different splits of the original sample material, in which case the statistical comparison will reflect

a combination of intra-laboratory analytical errors as well as the errors due to sample preparation.

- The paired values can be analyses performed by the same laboratory of two separate field samples that were very closely spaced, in which case the statistical comparison will reflect a combination of intra-laboratory analytical errors, sample preparation errors, sample collection errors and genuine short scale variations.

There are many ways that different samples, different splits of the same sample, different laboratories and different analytical techniques can be combined to provide pairs of experimental values. The interpretation of the paired values that result from such experiments is always easier if the QA/QC program is designed to isolate as much as possible the different factors that contribute to total sampling error.

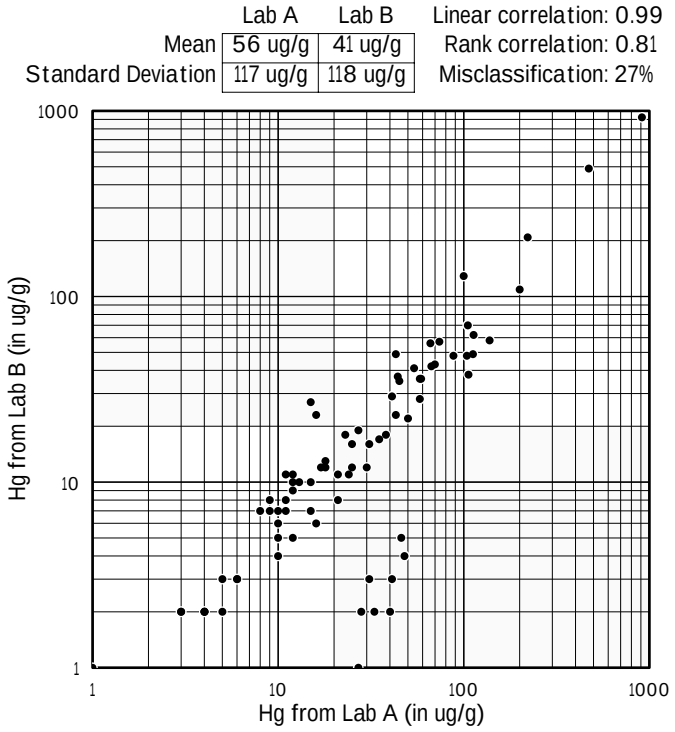

Figure 4 Comparison of analyses from different laboratories.

Figure 4 shows a scatterplot and some summary statistics for the mercury analyses that two laboratories produced for subsamples that were created by splitting the sample material in the field. This particular example reveals a systematic bias; the paired values tend to plot slightly off the main diagonal and the mean of the values reported by the two labs is noticeably different. The scatterplot also reveals that one of the labs may have a problem with inadvertent shifts of the decimal place; there is a handful of points along the bottom of the plot that would be more consistent if the Lab A value was lower by a factor of 10 or the Lab B value was higher by a factor of 10.

The example in Figure 4 also shows the advantage of reporting both the linear and the rank correlation coefficients. The strong skewness of the data makes the linear correlation coefficient quite sensitive to the extreme values; the rank correlation coefficient, a more stable statistic, shows that the high linear correlation coefficient in this example is due largely to the fact that the two labs were in very good agreement for the very highest few pairs of sample values.

Figure 4 reports the percentage of samples for which the labs disagreed on the classification for a remedial action threshold of 20 ug/g. For 21 of the 78 pairs, the two labs disagreed on whether the sample was contaminated; these 21 pairs plot in the shaded regions of the scatterplot. The 19 pairs that plot in the shaded region on the lower right were deemed contaminated by Lab A but not by Lab B; the 2 pairs in the upper left were deemed contaminated by Lab B but not by Lab A.

One of the shortcomings of a statistical analysis of paired observations is the ambiguity about which set of data is more reliable. Though the statistical summary of the mercury data shown in Figure 4 definitely reveals some problems, it is not clear if the problems lie with the analyses from Lab A or those from Lab B (or both). The best way to resolve such ambiguity is with control charts that directly address the accuracy and precision of each set of the paired sample values.

Another shortcoming of a statistical analysis of paired observations is that it may not reveal a systematic bias. Even if paired analyses show a strong agreement, this does not necessarily mean that both sets of values are accurate; it is possible that both sets of values share a common systematic bias.

Blank samples

QA/QC studies of data from contaminated sites need to pay particularly close attention to the possibilities of external contamination and cross-contamination between samples. With some of the contaminants being measured in trace quantities, external contamination can create considerable confusion in remediation planning if the materials used to collect, store and transport the samples are introducing measurable quantities of the contaminants of concern. Cross-contamination between samples can also create difficulties for remediation planning if material from an uncontaminated area becomes contaminated by material from elsewhere on the site.

Material that is known to be free of contamination can be inserted in the sampling procedure to provide experimental evidence of contamination. Such samples are usually called "blanks" and can be used to monitor contamination at various stages in the entire sampling procedure. The design of a program involving blank samples needs to consider all of the possible sources of contamination and all of the pathways for cross-contamination; without appropriate blank samples at each step, it may be difficult to interpret a finding of contamination and to develop an improved procedure that avoids the contamination. For example, trace amounts of chromium can be introduced by a variety of sources. Soil samples could be cross-contaminated by the chromium from refractory bricks if they are stored in the same area; chromium can also be introduced into samples by various metallic instruments. If blank samples prepared in the field reveal measurable increases in the chromium content, we may not know exactly where the trace amounts of chromium are coming from unless we have separate sets of blank samples that allow us to distinguish chromium cross-contamination during storage at the site from external chromium contamination introduced by metallic instruments in the lab.

OTHER ASPECTS OF QA/QC

QA/QC should not focus exclusively on the errors that affect the sample values. Statistical studies often depend on other quantitative information, such as sample location, and often also make use of qualitative information, such as descriptive logs of soil lithology. These other types of information often call for different QA/QC practices and procedures than those used for the sample values; the guiding principles, however, remain the same: we want to know the reliability of the information, we want to detect and correct spurious values and we want to minimize the impact of errors on the conclusions of our study.

Location information

Sample location errors can be minimized through careful surveying practices. Whenever possible, a standard reference coordinate system, such as UTM coordinates, should be used to record sample locations. If local coordinates are used, the procedure for converting these to a standard reference system should be documented. This can usually be accomplished by documenting the UTM coordinates of the origin of the local grid as well as any rotation between local north and UTM north.

If the only record of sample locations is a map, the sample locations will become increasingly unreliable as they are transcribed onto other maps. After several generations of copying and remeasuring, the original and correct sample locations are often so poorly known that the sample information is useless for location remediation planning. To prevent such problems, the coordinates of every sample location should be tabulated so that others can refer directly to the exact coordinates rather than trying to pick them off a copy of a map.

Qualitative information

Descriptive information, such as soil lithology, has a large component of subjectivity; the colour and texture that one person uses to describe a soil sample is often not the same as those that another person would use. As soon as it becomes apparent that descriptive information needs to be recorded for a particular site, we should standardize the collection of this information by preparing a form on which descriptive information can be recorded. When several people are collecting descriptive information, there will be more coherence between their descriptions if they are all given a standard set of reference materials, such as colour charts or grain size diagrams, that help them to calibrate their subjective visual judgement.

A complete photographic log of the samples is a very useful supplement for descriptive information and can be created using high quality film with relatively inexpensive photographic equipment. Variations in lighting conditions can be monitored and controlled by including a standard colour chart on every photograph. The existence of such a photographic record is often invaluable when old samples need to be relogged for descriptive information that has not yet been recorded because its importance was not initially recognized.

Merging data bases from different sources

In large contaminated site studies, it is common to find that the available data were gathered in different sampling campaigns by different organizations. When data from different sources are merged into a single data base, it is important to maintain a record of the original source for each piece of information. Long after the data have been merged, a statistcal QA/QC study may detect that certain data are less reliable than others. For example, one of the organizations responsible for sampling may choose to use larger boreholes than those used by another organization that has also collected borehole samples; since the size of the sample may affect the reliability of subsequent analyses, it may eventually be important to be able to distinguish the information generated by one organization from that generated by another. Similar concerns arise with location information when different organizations use different surveying practices, and with qualitative information when different organizations have different levels of expertise in recognizing and describing geological and geotechnical properties of the soil.

RECOMMENDED PRACTICE

1. A statistical study of contaminated site data should be accompanied by documentation of the reliability of any data that are critical to the study's conclusions.
2. The entire sampling procedure, including the collection, preparation and analysis of the sample, should not impart any systematic bias. For large studies in which more than 100 samples will be collected and analyzed, control charts should be used to monitor and control the accuracy and precision of the analyses. A t-test should be used to determine whether the average of repeat analyses is significantly different from the established reference value.
3. Sample precision should be monitored through control charts and through paired analyses of separate splits of the same sample material. For sample material that is split in the field, the paired analyses of the separate sample measurements should show a rank and linear correlation of 0.95 or greater for metallic and inorganic contaminants, and 0.90 or greater for organic contaminants.
4. Whenever QA/QC reveals a significant systematic bias or an unacceptably high imprecision, specific corrective action should taken and the results documented.

REFERENCES AND FURTHER READING

The guidance document entitled *UNIVARIATE DESCRIPTION* provides information on the summary statistics used in this document and also presents a more detailed discussion of data base compilation and verification. *BIVARIATE DESCRIPTION* provides information on scatterplots and their summary statistics. In addition to the other guidance documents in this series, the following references provide useful supplementary material.

Cochran, W.G., *Sampling Techniques*, 3rd edition, John Wiley & Sons, New York, 1977.

Merks, J., *Sampling and Weighing of Bulk Solids*, Series on Bulk Materials Handling, Volume 4, Trans Tech Publications, Clausthal, Germany, 1985.

Pitard, F., *Pierre Gy's Sampling Theory and Sampling Practice*, Volumes I and II, CRC Press, Boca Raton, Florida, 1989.