(5) 73770

Comparing Sites to Background

Part 1

A comparison of site measurements to background levels of a contaminant of concern (COC) often is employed when background measurements for the COC exceed a risk-based standard. In a site investigation and cleanup conducted under either RCRA or CERCLA, background data should be used to: reduce a list of contaminants of potential concern (COPC) to a shorter list of COPCs in a remedial investigation (RI) or a baseline risk assessment; establish remedial goals or cleanup levels; and verify cleanup to those levels.

Unfortunately, no de facto method is available to compare site measurements to background measurements and to test if site concentrations for the COC are above background levels. Typically, the statements "site concentrations are no different than background for a specific contaminant" or "reduce the concentration of the COC to background levels" are easy to understand: we desire that the site has no more COC contamination than already exists in background. However, this is difficult to determine using quantitative statistical methods, since one must decide the metric that will be used to compare the site and background concentrations. Different metrics assume different underlying hypotheses and pose different decision error criteria to meet data quality objectives (DQOs).[1]

METHODOLOGY

To compare site concentrations to background concentrations for a contaminant, three common metrics are used and discussed in the literature:

1. Comparing the maximum observed COC concentration in the site to the maximum observed COC concentration in background.
2. Comparing the mean COC concentration in the site to the mean COC concentration in background.
3. Comparing the site COC concentration distribution with the background distribution to look for a shift (to the right) of the site distribution.

This article will discuss the first metric; the latter two will be covered in part two of this article. Our discussion of these metrics focuses on the ability of each to keep decision error rates low, under the assumption that the site and background areas have similar COC concentration distributions.

COMPARING MAXIMA

Comparing the maximum site COPC concentration measurement to the maximum contaminant concentration in the background data set is a common practice when reducing COPCs to COCs in a remedial investigation for inclusion in a baseline risk assessment. In this setting the methodology is: if one or more site COPC concentration measurement exceeds the maximum background measurement, then the COPC is considered a COC and must be carried through the risk assessment. Of course, this approach adds unnecessary cost to the remediation effort if the COPC is actually at background levels. Also, this methodology has been used in some instances to determine whether a site has been remediated to background levels. Two key issues make this methodology unsuitable for remedial work.

1. If an equal number of samples are measured for the contaminant in both the site and background, and if the site and background truly have the same COC concentration distribution, then the probability is 0.5 (50%) that the maximum measurement occurs in either the site or background unit. Thus, the chance is 50% that a COPC will be declared a COC, when in fact the site is actually at background levels.
2. If unequal numbers of sample observations are measured in the site and background, and if the site and background units have the same COC

By Robert F. O'Brien and Richard O. Gilbert

concentration distributions, then the probability is higher that the unit with the most observations has the maximum observed COC concentration. If the site has m observations and the background unit n observations, then the probability is m/(n+m) that the site will have the largest observation. For example, under the assumption that the site and background have identical COC concentration distributions, if the site has 10 observations and the background unit 20 observations, then the probability is 10/(10+20) =1/3 that the site has the maximum concentration. If the site has more observations than the background unit, then the probability that a site measurement exceeds the largest background measurement is >50% (since m/(n+m)>0.5 in this case).

The probability that the maximum occurs in the site, under the assumption that both the site and background units have the same distribution, depends on the number of observations in the site and in the background unit. The probability of making a decision error in declaring the COPC a COC or in declaring a site contaminated is usually large, unless m/(n+m) is small. This is seldom the case in practice, for to get a constant decision error rate in the range of approximately 5%, the ratio of site measurements to background measurements would have to be on the order of 1 to 20.

***Editor's note:** Part 2 of this article, which will be published in an upcoming issue, will discuss the statistical approaches of comparing means and comparing distributions. It also will make recommendations on which contamination scenarios and statistical tests are most compatible in terms of reducing decision error.*

***Richard F. O'Brien** is with Pacific Northwest National Laboratory in Richland, WA.*

***Richard O. Gilbert** is with the Washington, DC, office of PNNL.*

REFERENCE

1. EPA. 1994a. Guidance for the Data Quality Objectives Process. EPA QA/G-4. U.S. Environmental Protection Agency, Quality Assurance Management Staff, Washington, DC.

Comparing Sites to Background, Part 2

By Robert F. O'Brien and Richard O. Gilbert

Part 1 of this column, which appeared in the September/October 1997 issue of ENVIRONMENTAL TESTING & ANALYSIS, discussed methodology and maxima comparison. In this issue, the authors complete their overview by discussing the statistical approaches of comparing means and distributions and making recommendations on which contamination scenarios and statistical tests are most compatible to reduce decision error.

Comparing Means

Comparing the mean site concentration to the mean background concentration is a common method that is followed in Resource Conservation and Recovery Act (RCRA) and Comprehensive Environmental Response, Compensation and Liability Act (CERCLA), or Superfund, remediation efforts. Under the assumption that the site and background unit have the same true mean and concentration distribution, common parametric tests for equality of means are the t-test (for normal and lognormal data) and other methods as discussed in Lehmann.[1]

The null and alternative hypotheses for this approach are:

(1) $H_0: \mu_S - \mu_B = 0$
$H_A: \mu_S - \mu_B \geq 0,$

where μ_S and μ_B are the site and background means respectively. H_0 is the null hypotheses assumption of the equality of the true site and background means, and H_A is the alternative hypothesis that the site mean is greater than the background mean..

An advantage of this test of means is that it can be easily designed to produce acceptably low decision error rates, even for small or moderate numbers of samples. However, for large numbers of sample observations, a drawback in testing for the equality of the means is that a small difference in the means will almost always be detected. This raises a question. While the test indicates that the site and background means are not equal, does a small increase in the site mean pose any real additional human or ecological risk? For small differences between site and background means, the additional risk posed by the elevated site mean may be negligible. Thus, the real question should be: Is mean concentration of the contaminant in the site greater than D units above background, where D is a concentration that poses an unacceptable added risk above the mean background concentration?

In this form, the null and alternative hypotheses would be:

(2) $H_0: \mu_S - \mu_B \leq D$
$H_A: \mu_S - \mu_B > D,$

or equivalently,

(3) $H_0: \mu_S \leq c\,\mu_B$
$H_A: \mu_S > c\,\mu_B,$

where c is a real number greater than 1.

For normal and lognormally distributed data, t-tests are easily developed to test hypotheses (2). For distributions such as the gamma a test for (3) has been developed by O'Brien, Sinha and Smith.[2] A somewhat different approach to testing in these situations is the use of tests for "bio-equivalence."[3]

Another common problem "t" in testing means is that under the null hypothesis the site and background units should have the same contamination distribution. For example, in Figure 1, X is the site distribution and Y is the background distribution. It is easily seen that they have different distribution shapes though they have the same mean of 3. In situations like this, the common t-test and other parametric tests are inappropriate, since they assume the underlying distributions have the same shape. In these situations, however, it is still possible to carry out tests of the abovementioned hypotheses using bootstrapping techniques.[4]

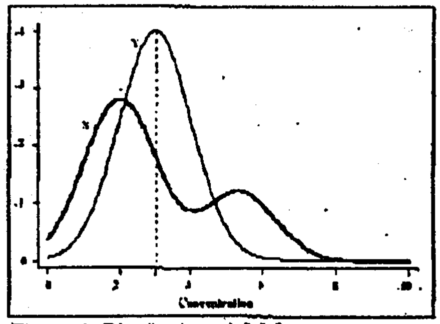


Figure 1. Distribution of COC concentration for site, X, and background, Y, where X and Y have same mean but different shapes.

Comparing Distributions

When it is difficult to decide which

probability distribution is appropriate for the site and background measurements, non-parametric tests, such as the Wilcoxon Rank Sum (WRS) test and the quantile test, are suggested.[5] These methods are also appropriate if there is a moderate amount of left-censoring in the concentration data, due to observations that were recorded as falling below the laboratory quantitation limit. The WRS test is appropriate when the central tendency of the site distribution is shifted to the right of the background distribution. The quantile test is appropriate when the right tail of the site distribution is shifted towards the right of the right tail of the background distribution.

The benefit of the WRS and quantile tests is that they can achieve specified small decision error rates.

The benefit of the WRS and quantile tests is that they can achieve, with a moderate number of samples, specified small decision error rates (for falsely rejecting the equivalence of the site and background distributions). However, like the tests of means, the WRS test can lead to erroneous rejection of the equivalency of two distributions that are negligibly different, if there are a large number of observations in the site and background.

The hypotheses that are tested in the WRS test are:

H_0: $P(X > Y) = 1/2$ against
H_A: $P(X > Y) > 1/2$,

where $P(X > Y)$ is the probability that a concentration measurement observed in the site is greater than a measurement from the background unit.

For the quantile test, the hypotheses tested are:

H_0: $p = 0$, $D/\sigma = 0$ against
H_A: $p > 0$, $D/\sigma > 0$,

where p is the percentage of the site measurements that are above background, and D/σ is the amount (in units of standard deviation) that p percent of the site measurements are shifted to the right of the background, even when the difference between the two underlying distributions is small enough to pose no increase in risk.

In Figure 2, X represents the background distribution of the contaminant concentration, and Y and Z the concentration distributions of two different sites.

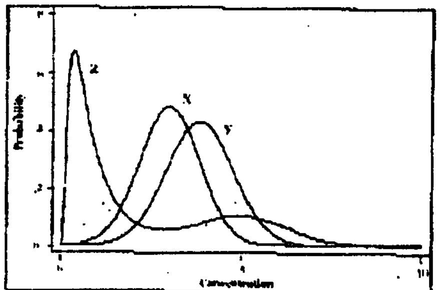


Figure 2. Contamination distributions of three different shapes.

In this case, the WRS does a better job of distinguishing between X and Y than X and Z, while the quantile test does a better job of distinguishing between X and Z than X and Y. However, a test that provides the best of both procedures, a combined "tandem" WRS and quantile test, is discussed in Hardin and Gilbert.[6] In most situations, this "tandem" test actually performs better than either test separately.

One drawback of the WRS and quantile tests is that, as in the test for means, if a large number of samples is taken in both the site and background units small differences may be detected by the tests. However, we can adjust these tests to handle incremental concentration differences or shifts that are believed to add little risk.

Another alternative in non-parametric testing is to test for a difference between the proportion of the site measurements and the proportion of background measurements that are above the risk-based standard for the contaminant of concern (COC). This reduces to a binomial testing situation, which can also incorporate the condition that the difference between the proportions has to be greater than some specified amount D that would pose an increase in risk.

Conclusion

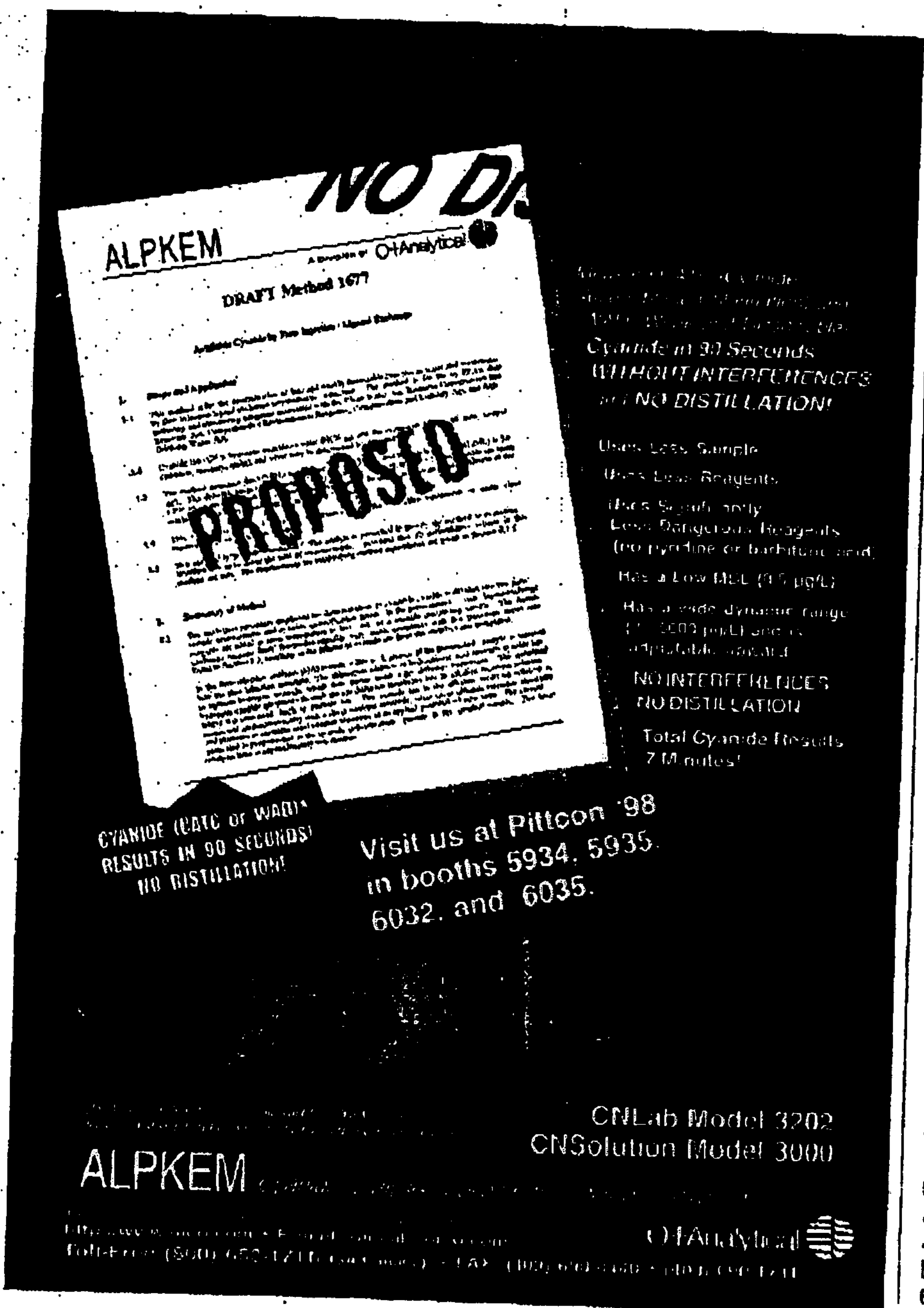
Simply comparing the maximum site concentration to the maximum background concentration measurement, as discussed in Part 1, offers inadequate protection against a declaration that site is elevated above background when in fact, the site is no different than background. Therefore, this approach should be avoided unless a ratio of approximately 20:1 background to site measurements can be attained.

If the comparison of means is chosen to verify whether background standards are attained testing should be carried out to determine whether or not the site mean is more than D units above the background mean, where D units represent a significant increase to the risk of human or ecological exposure. This approach supports reasonable decision error rates that will protect against erroneously declaring the site and background as different from one another, when in fact, the site and background pose similar contaminant exposure risks.

If one decides to test for a shift in the site distribution to the right of the background distribution, the WRS test combined with the quantile test should be employed. This offers reasonable decision error rates, is fairly easy to carry out, and is robust in many conditions. These nonparametric methods can handle cases in which the site and background contaminant concentration distributions are not identical, and where left-censored observations impede the use of parametric methods.

Robert F. O'Brien is with Pacific Northwest National Laboratory (PNNL) in Richland, WA.

Richard O. Gilbert is with the Washington, DC, office of PNNL.



References

1. Lehmann, E.L. *Testing Statistical Hypotheses.* John Wiley, New York. 1986.

2. O'Brien, R.F., B.K. Sinha, and W.P. Smith. A Statistical Procedure to Evaluate Cleanup Standards. *Journal of Chemometrics.* 5 (3), pp. 249-261. 1991.

3. McDonald, L.L. and W.P. Erickson. Testing for the bioequivalence in field studies: Can a disturbed site be adequately reclaimed? *Statistics in Ecology and Environmental Monitoring,* pp. 183-197. O.J. Fletcher and B.F. Manly, eds. University of Otago Press, New Zealand. 1994.

4. Efron, B. and R.J. Tibshirani. *An Introduction to the Bootstrap.* Chapman & Hall, New York. 1993.

5. *Statistical Methods for Attainment of Cleanup Standards Vol 3: Reference Based Standards for Soils and Solid Media.* EPA-230-R-94-004. U.S. EPA, Washington, D.C. 1994b.

6. Hardin, J.W. and R.O. Gilbert. *Comparing Statistical Tests for Detecting Soil Contamination Greater than Background.* PNL-8989. Pacific Northwest Laboratory, Richland, WA. 1993.