Comparing Background and Site Chemical Concentrations

Steve Roberts, UF
Ligia Mora-Applegate, FDEP

Determining the Need for Background Sampling

Background samples are needed when chemical concentrations at a site exceed risk-based cleanup levels, yet may represent natural background conditions.

Basic Approaches

- Non-statistical
  - Compare site max with 2x the average of the background samples, or the maximum background concentration (whichever is lower)
  - Works with limited number of samples
  - Conservative
  - Well accepted; in Region 4 guidance
- Statistical
  - Stronger approach for comparison
  - Requires more background data than are available for most sites

"Twice the mean" approach

- Twice the mean and maximum background concentration are intended to represent the upper end of the range of background concentrations.
- In theory, could calculate 2x the mean with as few as two background samples, although more is obviously better.
- If chemical passes this test, it can be excluded as a chemical of concern.
- Region 4 accepts this approach for soil, but not groundwater.

Statistical Approach

- Guidance is available from USEPA
  - "Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites" (EPA 540-R-01-003, 2002)
- A number of decisions must be made:
  - Test Form (hypothesis to be tested)
  - Statistical test
  - Confidence level
  - Power
  - "Substantial difference" to be detected

Hypothesis Testing

Background Test Form 1:

- The null hypothesis (H_o) states that the mean chemical concentration of site samples is not different from the mean concentration in background samples.
- Baseline assumption is that the chemical is not a contaminant; concentrations reflect background conditions.
- “innocent until proven guilty”

Hypothesis Testing

Background Test Form 2:

- The null hypothesis (H_o) states that the mean chemical concentration of site samples is different from background by more than a specified concentration level.
- Baseline assumption is that concentrations on site reflect contamination.
- “guilty until proven innocent”

Practical considerations

- Choice depends in part on which is more important to avoid:
  - Deciding a site is contaminated when it is not.
  - Deciding a site is not contaminated when it is.
- Form 1 benefits the PRP if the test is under powered; Form 2 provides incentive to collect more samples.
- Form 2 requires the site manager to decide what constitutes a substantial difference from background -- “how close is close enough?”

Statistical tests

- Parametric (e.g., t-test)
  - Best used with normally distributed data
  - Best used with large data sets (>25 samples) with few non-detects
- Non-parametric (e.g., WRS)
  - No requirement for normally distributed data
  - Works well with non-detects
  - Useful for smaller data sets (>15-20 samples)

Confidence and Power

- Type I error: Rejection of the null hypothesis when in fact it is true. The probability of making a Type I is error is α.
- Type II error: Failure to reject the null hypothesis when in fact it is false. The probability of making a Type II error is β.
- Confidence = $(1-\alpha)$; Power = $(1-\beta)$.
  - Need to find the right balance between confidence, power, and sample size.
  - USEPA guidance has suggestions for confidence and power choices.

Sample Size

- More is better
- Generally, 15-20 samples [each for background and for the site] is minimum
- Exact number of samples needed can be calculated.
  - Need to specify the confidence, the power, and the size difference you want to be able to detect.
  - Results of a preliminary statistical comparison can be informative ("close -- if we had 20 more samples that looked like these, the results would be statistically significant")

"Substantial Difference"

- Form 2 requires you to specify a "substantial difference" (*S*)
  - Basically, *S* is how close to background is close enough.
  - *S* can't be smaller than what your statistical test can detect.
  - Lots of ways to come up with an *S* value, but it is ultimately a management call.

Outliers

- Outliers are measurements that are significantly larger or smaller than the rest of the data -- they don't belong to the same population.
- When used in descriptive statistics they can incorrectly estimate the true value of the population.
- Parametric tests are especially sensitive to outliers.
- Tests are available to identify outliers; subjective input is sometimes important.

Non-Detects

- Parametric statistical tests require values for all data points. Non-detects cannot be used in these tests, so surrogate values are assigned. Options for surrogate values include:
  - Half the detection limit
  - Detection limit (L)/$2^{1/2}$
  - A random value between 0 and the detection limit
- Nonparametric tests do not require surrogate values for non-detects. In these tests, a rank is assigned to each value, and it is the rank which is used for the test.

Example

Background Test Form 1

Data (mg/kg)	Source
ND	Background
ND	Background
ND	Background
3.5	Background
2.2	Background
2.9	Background
2.5	Background
3.4	Background
2.6	Background
3.5	Background
2.8	Background
2.4	Background
3.3	Background
1.9	Background
2.9	Background

Data (mg/kg)	Source
3.4	Site
3.6	Site
2.4	Site
3.1	Site
5.0	Site
2.2	Site
3.4	Site
1.8	Site
4.0	Site
3.7	Site
2.2	Site
4.1	Site
2.8	Site
3.2	Site
3.3	Site

Background Test Form 1

		Site	Background
ND	Background		2
ND	Background		2
ND	Background		2
1.8	Site	4	
1.9	Background		5
2.2	Background		7
2.2	Site	7	
2.2	Site	7	
2.4	Background		9.5
2.4	Site	9.5	
2.5	Background		11
2.6	Background		12
2.8	Background		13.5
2.8	Site	13.5	
2.9	Background		15.5
2.9	Background		15.5
3.1	Site	17	
3.2	Site	18	
3.3	Background		19.5
3.3	Site	19.5	
3.4	Background		22
3.4	Site	22	
3.4	Site	22	
3.5	Background		24.5
3.5	Background		24.5
3.6	Site	26	
3.7	Site	27	
4.0	Site	28	
4.1	Site	29	
5.0	Site	30	
		279.5	185.5

- Step 1 - Combine the background and site data sets in ascending order (treat non-detects as zero values).
- Step 2 - Assign rank to the values using an average rank for equal values.
- Step 3 - Add the ranks for the background and site data separately.
- Step 4 - Compare the site value with the WRS test statistic (see handout sheet; use second number because it is a two-tailed test).

Background Test Form 1

		Site	Background
ND	Background		2
ND	Background		2
ND	Background		2
1.8	Site	4	
1.9	Background		5
2.2	Background		7
2.2	Site	7	
2.2	Site	7	
2.4	Background		9.5
2.4	Site	9.5	
2.5	Background		11
2.6	Background		12
2.8	Background		13.5
2.8	Site	13.5	
2.9	Background		15.5
2.9	Background		15.5
3.1	Site	17	
3.2	Site	18	
3.3	Background		19.5
3.3	Site	19.5	
3.4	Background		22
3.4	Site	22	
3.4	Site	22	
3.5	Background		24.5
3.5	Background		24.5
3.6	Site	26	
3.7	Site	27	
4.0	Site	28	
4.1	Site	29	
5.0	Site	30	
		279.5	185.5

- The WRS test statistic for background = site = 15 and $\alpha = 0.05$ is 273.
- Since 279.5 > 273, we reject the null hypothesis (site concentrations are not significantly greater than background) at the 95% confidence level.

Background Test Form 2

Data (mg/kg)	Source
ND	Background
ND	Background
ND	Background
3.5	Background
2.2	Background
2.9	Background
2.5	Background
3.4	Background
2.6	Background
3.5	Background
2.8	Background
2.4	Background
3.3	Background
1.9	Background
2.9	Background

Data (mg/kg)	Source
3.4	Site
3.6	Site
2.4	Site
3.1	Site
5.0	Site
2.2	Site
3.4	Site
1.8	Site
4.0	Site
3.7	Site
2.2	Site
4.1	Site
2.8	Site
3.2	Site
3.3	Site

- In Background Test Form 2, background values are adjusted by adding a substantial difference to each value. In this example, the standard deviation was chosen as the substantial difference.

Background Test Form 2

		Site	Background
0.7	Background		2
0.7	Background		2
0.7	Background		2
1.8	Site	4	
2.2	Site	5.5	
2.2	Site	5.5	
2.4	Site	7	
2.6	Background		8
2.8	Site	9	
2.9	Background		10
3.1	Background		11.5
3.1	Site	11.5	
3.2	Background		13.5
3.2	Site	13.5	
3.3	Background		15.5
3.3	Site	15.5	
3.4	Site	17.5	
3.4	Site	17.5	
3.5	Background		19
3.6	Background		21
3.6	Background		21
3.6	Site	21	
3.7	Site	23	
4.0	Background		24.5
4.0	Site	24.5	
4.1	Background		26.5
4.1	Site	26.5	
4.2	Background		28.5
4.2	Background		28.5
5.0	Site	30	
		231.5	233.5

- The two sample populations are then combined and ranked as in Background Test Form 1.
- For this test, the sum of the background ranks is the test statistic.

Background Test Form 2

		Site	Background
0.7	Background		2
0.7	Background		2
0.7	Background		2
1.8	Site	4	
2.2	Site	5.5	
2.2	Site	5.5	
2.4	Site	7	
2.6	Background		8
2.8	Site	9	
2.9	Background		10
3.1	Background		11.5
3.1	Site	11.5	
3.2	Background		13.5
3.2	Site	13.5	
3.3	Background		15.5
3.3	Site	15.5	
3.4	Site	17.5	
3.4	Site	17.5	
3.5	Background		19
3.6	Background		21
3.6	Background		21
3.6	Site	21	
3.7	Site	23	
4.0	Background		24.5
4.0	Site	24.5	
4.1	Background		26.5
4.1	Site	26.5	
4.2	Background		28.5
4.2	Background		28.5
5.0	Site	30	
		231.5	233.5

- The WRS test statistic for background = site = 15 and $\alpha = 0.05$ is 273.
- Since 233.5 < 273, we do not reject the null hypothesis (site concentrations are significantly greater than background) at the 95% confidence level.

Background Test Forms

Both Background Test Forms result in the same conclusion - the site values are significantly greater than background values.

Not Recommended

- Comparing the average (or 95% UCL) concentration on site with twice the background.
  - Risks from soils at the site could be up to twice background levels and still pass this test.
- Comparing the average (or 95% UCL) on site with the average concentration of background.
  - This test is highly influenced by the number of samples taken and their location.
  - It is possible to pass this test even with a few obviously contaminated areas.
  - It is also possible to fail the test with site concentrations that fall well within the background range.

Other States' Approaches

- Alabama - Compare site maximum with 2x arithmetic mean of the background concentrations.
  - 4 samples minimum required.
- Arkansas - Use EPA "Guidance for Comparing Background …" and Region 8 guidance.
  - In some situations have compared site concentrations with background mean + 2SD; 4 samples minimum

Other States' Approaches

- Kentucky - Use three criteria:
  - For inorganics, the mean concentration on site must be less than the 95% UCL of the mean of background.
  - At least half of the data points should be less than the 60th percentile of background.
  - No data points should be above the 95% percentile.

Other States' Approaches

- New Jersey
  - Maximum concentration on site cannot exceed maximum background concentration.
  - A minimum of 10 background samples are collected. Examples of excluded locations for background samples are provided, e.g., parking lots, areas near railroad tracks, etc.
  - Formal criteria for identifying and excluding outliers are provided.

Other States' Approaches

- Utah - No definitive guidance; different approaches have been used, e.g., maximum on site compared with mean + 2SD background concentration.
- Ohio - Maximum concentration on site cannot exceed background levels.
- Wisconsin - No definitive guidance; a statistical approach can be proposed if enough data are available.