THE USE OF CONFIDENCE LIMITS TO ASSESS ASBESTOS IN SOIL CONCENTRATIONS

Chris Hillman, Tonkin & Taylor Ltd[1]

Introduction

A key element to assessing the management requirements for contaminated land is to compare soil test data against published human health criteria. For asbestos in soil, assessment criteria are provided by the New Zealand Guidelines for Assessing and Managing Asbestos in Soils (BRANZ, 2017).

Soil testing datasets frequently include a number of results that exceed a particular guideline value. This can be addressed by estimating Upper Confidence Limits (UCLs), usually around the mean. However, there is a reluctance to estimate UCLs for asbestos data in the contaminated land industry because asbestos is not seen to "behave like other contaminants": Asbestos in soil data usually has a skewed distribution (a small number of very high results, as shown in Figure 1) along with a high number of non-detects (typically more than 50% of the dataset).

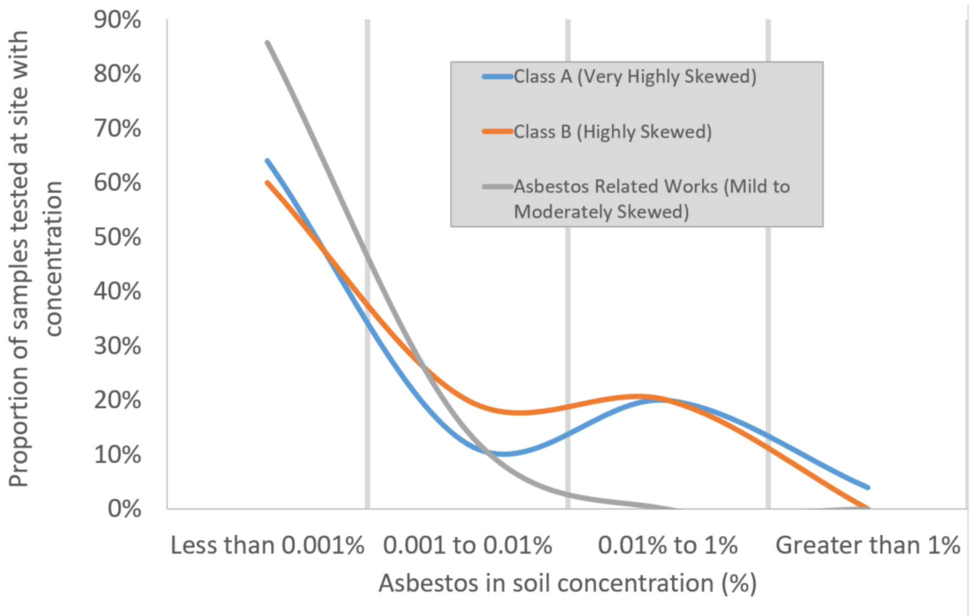

Figure 1: Distribution of asbestos concentrations from laboratory analysis (Skewness after Department of Energy and Environmental Protection, 2014)

[1] Contact: CHillman@tonkintaylor.co.nz

There is no specific guidance on applying confidence limits to asbestos in soil data. This paper provides a roadmap of how UCLs can be estimated for asbestos in soil using the following standard iterative steps:

- Outliers identification
- Calculate the UCL
- Compare the UCL with the guideline value

This paper assumes that the reader is familiar with the application of these steps to chemical contaminant datasets and with the BRANZ guidelines (BRANZ, 2017).

A glossary of terms in included as Appendix A.

Sufficient numbers of samples

The Ministry for the Environment (MfE, 2011) recommends a minimum of 30 samples to provide a robust understanding of the nature of a sample population. Sampling data from USEPA Superfund sites has shown that data sets with fewer than 10 samples per stratum provide poor UCLs, with 10 being the minimum for reasonably reliable estimates. 10 to 20 samples provided somewhat better estimates (for normally distributed to slightly skewed data). 20 to 30 samples provide a fairly consistent estimate (USEPA, 1992 & 2013; ITTRC, 2012).

Larger samples are required for moderately skewed data to effectively identify outliers and estimate a meaningful UCL (USEPA, 2013). Highly skewed data requires more samples still. Research undertaken by the USEPA (USEPA, 2006b) suggests that the need for high sample numbers can be mitigated by increasing the coverage of the UCL (e.g. move from a 95% UCL to 97.5% or 99% UCL for comparison with the guideline value) and by using specific methods to calculate the UCL (see below).

Sufficient volume of samples

Standard asbestos investigation sample volumes are 10L for field screening and 500mL for laboratory analysis. The basis for these volumes is unclear, however, based on the number of non-detects generally encountered when investigating asbestos in soils, they would appear too low to fully characterise each sample. To illustrate the problem, we can imagine a soil containing only cement boarding fragments of 30mm diameter; at the residential guideline for bonded ACM of 0.01% (BRANZ, 2017), there would be one 30mm diameter fragment for every 16 litres of soil[2], ie one fragment for every one and a half 10L samples collected. The lower the concentration of asbestos as a whole, the more dispersed the fragments and the more volume of sample needed to characterise the soil.

Pierre Gy undertook extensive work to enable assessment of the order of magnitude of sample volume needed to characterise heterogeneous materials (See Appendix B). The theorem is much more involved than the simple example above and has been applied to field screening and laboratory analysis of asbestos in soil in Figure 2 below.

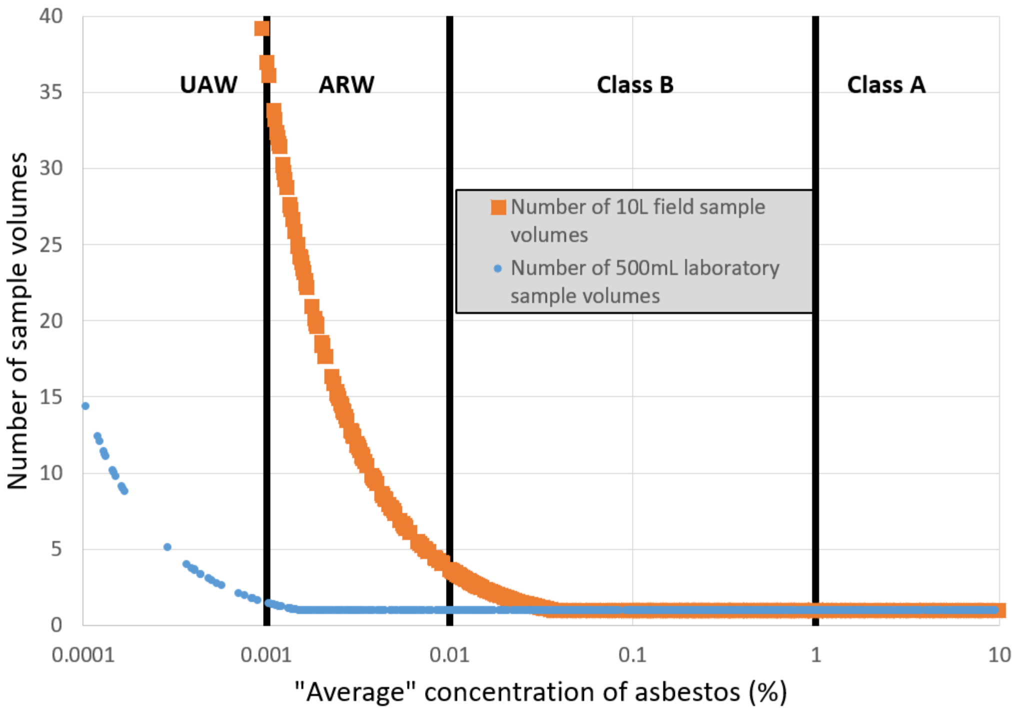

Figure 2: Order of magnitude of volumes of soil required to characterise asbestos concentrations (plot generated using a Monte Carlo Simulation for a range of asbestos concentrations, using the parameters described in Appendix B)

The applicability of Gy's theorem has been criticised by some (although an alternative has not been suggested) and is uncalibrated for asbestos in soils. However, based on Figure 2,

---

[2] Assume fragments are 30mm diameter, which would have an area of $7cm^2$ and, at 0.7cm thick, a volume of $4.8x10^{-6}m^3$. Mass of asbestos = $3000kg/m^3$, with 20% asbestos in the fragments. Therefore mass of asbestos in a fragment = 3000 x 100/20 x $4.8x10^{-3}m^3$ = 0.003kg. Volume of soil for this fragment to comprise 0.01% = 0.003 x (100/0.01) = 30kg (16.5 litres at $1800kg/m^3$ soil density)

10L and 500mL samples appear appropriate for assessing soil concentrations down to approximately 0.015% and 0.001% respectively. For concentrations lower than this, the volume required to characterise the soil increases markedly (i.e. asbestos could be present in the soil, but not included in the sample collected).

This indicates why there are a significant number of non-detects associated with asbestos in soil investigations and that just because a result comes back as non-detect does not mean that there is no asbestos present it he soil that surrounded the sample.

Outliers assessment

Outliers can distort conclusions drawn from the data and should be removed, but only where they can be fully explained.

It can be difficult to identify whether or not the small number of very high results typical of asbestos in soil datasets are outliers. As with all environmental datasets, to definitively identify outliers, the nature of the underlying population needs to be well understood. “Well understood” means:

- A robust desk study
- Sufficient samples

The standard approach to assessing outliers in data with non-detects is to replace the non-detects with their respective detection limits (or half the detection limit) before applying any statistical test. However, general consensus appears to be that substituting non-detects gives biased and unreliable results and should not be used. When the number of non-detects is large, it is often simpler to just ignore them (USEPA, 2013).

The process for identifying outliers is to first examine the data through graphical methods (e.g. a box plot or Q-Q plot) and by reference to the ground model (see Appendix C for examples of outlier identification). However, with skewed data that has not been successfully normalised, graphical methods may still be unsatisfactory.

The USEPA provide guidance on statistical outlier identification methods (USEPA, 2006a), however the general advice is that with small datasets (less than 30), statistical analysis to identify outliers should only be undertaken as a last recourse (MfE 2011) and for skewed data this minimum number is likely to be much higher, making it impractical for most asbestos in soil investigations. So asbestos in soil investigations are likely to be limited to interrogation of the ground model for outlier identification.

Calculate the UCL

High proportions of data below the detection limit can make calculating a UCL challenging.

MfE advice is to not report a UCL where more than 25% of the data is below the detection limit (MfE, 2011). However, internationally there are validated methods to derive a UCL from environmental data, even where there are a high proportion non-detects.

The USEPA accepts a number of methods for assessing ground contamination data against guideline values (USEPA, 2006a). For datasets with 50% to 90% non-detects, the "Test for Proportions" is recommended by the USEPA. However, large numbers of samples are required to apply this test, making it likely impractical for all but the most extensive of ground contamination investigations. Other USEPA guidance (USEPA, 2006b) addresses the difficulties of high numbers of non-detects by applying a statistical method to compute the sample mean that does not rely on any knowledge of the data distribution, the Kaplan-Meier Estimation Method. However, where there is a single detection limit on the left hand side of the data, the Kaplan-Meier Estimation Method is essentially substitution, which is indicated to give biased and unreliable results.

An alternative is to use Robust Order Statistics (ROS), which can cope with skewed data containing significant proportions of non-detects (Helsel 2012). ROS replaces non-detects with a model distribution (e.g. lognormal). However, rather than hypothesise values for results below the detection limit, it would seem prudent to replace any detected results below the method detection limit with the actual measured result and only use ROS where no asbestos is recorded at all.

If non-detects represent more than 80% of the dataset, then analysis is likely to be beyond the non-statistician.

A UCL can be calculated around the mean estimated using ROS. The appropriate UCL methods are dependent on the sample numbers and skewness of the dataset. USEPA recommended methods are summarised in Table 1 below. As can be seen from the table, low sample numbers for moderately and highly skewed data are mitigated by increasing the percent coverage of the UCL.

Table 1: USEPA recommended UCL methods for high numbers of non-detects

Skewness[2]	Sample size	% Non detects	Method
less than 1.0	Min 8 to 10[3]	Greater than 40%	Student t
1.0 to 1.5	Min 8 to 10[3]	Greater than 40%	BCA
1.5 to 2.0	More than 40 to 50	Greater than 50%	95% Chebyshev
	Less than 40 to 50	Greater than 50%	97.5% Chebyshev
2.0 to 3.0	More than 50 to 60	Greater than 0%	
	Less than 50 to 60	Greater than 0%	99% Chebyshev

Notes:

1. This table is a simplification of Table 9-1 in USEPA, 2006b, for lognormal and non-parametric datasets, with the Kaplan-Meier method removed.
2. Skewness = Standard deviation of log-transformed dataset
3. USEPA, 2006b

More detailed descriptions of the above methods are provided in USEPA, 2006b. The latest version of ProUCL includes each of these methods. An example of UCL calculation is provided in Appendix D.

Comparison with guidelines

The MfE guidelines (MfE, 2011) state that as a rule of thumb, the site will be acceptable if the 95% UCL is at or below the guideline, provided no result is more than twice the guideline value.

On the basis of USEPA research (USEPA, 2006b), comparison of the 97.5% UCL or 99% UCL with the guideline value, using appropriate methods, would appear reasonable where sample numbers are insufficient to reasonably use a 95% UCL. However, as the coverage of the UCL increases, so the estimated value approaches the maximum result in the dataset (i.e. there is little difference between the 99% UCL value and the maximum value in the dataset). Therefore, where datasets are small and moderately to highly skewed, calculating an appropriate UCL is only likely to be of value where the maximum values are only just above a guideline value.

The basis of the MfE requirement for no result to be more than twice the guideline value is unclear. Whilst this may be a practical limit for chemical contamination data, if applied to skewed asbestos in soil data then for the majority of sites the precautions required would be controlled simply by the highest result recorded. Therefore, this rule of thumb is unlikely to be practical for asbestos in soil data.

The last words in comparing the UCL to a guideline value should be given to the ground model:

The results should always be assessed in the context of the site, proposed land use and DQOs, and be related to the known information about the site history, sources of contamination and pathways for migration and target receptors. The basis for the derivation of any guideline should be understood and the suitability for use considered in the context of the site. (MfE, 2011).

Summary

- Asbestos in soil investigations generally provide mildly to very highly skewed datasets and upwards of 50% non-detects.
- To identify outliers, interrogation of the ground model is often the only practical method:
    - Graphical methods can be unsatisfactory if skewed data hasn't been successfully normalised.
    - For small datasets (less than 30), statistical analysis should only be undertaken as a last recourse.
- High proportions of data below the detection limit and skewness can make calculating a UCL challenging:
    - Robust Order Statistics (ROS) can be used to substitute non-detects of skewed data (up to about 80% non-detects) and Table 1 above lists subsequent UCL calculation methods that can be used, dependent on the skewness, proportion of non-detects and sample numbers.
    - High sample numbers are required to calculate a 95% UCL for moderately and highly skewed data. However, low sample numbers can be mitigated by increasing the percentage coverage of the UCL (e.g. use 99% UCL instead of 95% UCL).

- More than 80 to 90% of non-detects and calculation of a UCL is probably beyond the non-statistician (and may not be meaningful even then).

Conclusions

Confidence limits can be calculated for asbestos in soil data. However, this data often has a moderately to highly skewed distribution, along with low sample numbers.
Where this is the case, a UCL is only likely to be of value where the maximum values in the dataset are marginally above the guideline value.

Whether or not a UCL is calculated to be above or below a guideline value, statistical analysis of asbestos in soil data only provides one line of evidence to feed into the ground model. The ground model ultimately informs the precautions required to protect human health, not any one line of evidence.

Disclaimer

The opinions within this paper have been provided by the author to foster debate and should not be taken as those of Tonkin & Taylor.

Acknowledgements

Thank you to Shane Moore and Cath Kingston for their considered opinions (although they may not necessarily agree with everything in this paper). Also thank you to Linda Horvath for assistance obtaining the references used here and to Lorraine Barnes for her sharp reviewing eyes.

References

- BRANZ, Nov 2017, New Zealand Guidelines for Assessing and Managing Asbestos in Soil, BRANZ, Wellington
- Chartered Institute of Environmental Health & CL:AIRE, May 2008, Guidance on Comparing Soil Contamination Data with a Critical Concentration.
- Department of Energy and Environmental Protection, May 2014, Guidance for Calculating the 95% Upper Confidence Level for Demonstrating Compliance with the Remediation Standard Regulations
- Helsel, D.R., 2012, Statistics for Censored Environmental Data Using MINITAB and R, 2nd Edition, John Wiley & Sons

- Interstate Technology & Regulatory Council Incremental Sampling Methodology Team, Feb 2012, Incremental Sampling Methodology
- Interstate Technology & Regulatory Council, Dec 2013, 5.1 Graphical Methods, Accessed 5 August 2019, [https://www.itrcweb.org/gsmc-1/Content/GW%20Stats/5%20Methods%20in%20indiv%20Topics/5%201%20Grph%20Methods.htm#Box]
- Ministry for the Environment, revised 2011, Site Investigation and Analysis of Soils
- Minkkinen, P, (1987) Evaluation of the Fundamental Sampling Error in the Sampling of Particulate Solids, Analytica Chimica Acta, 196, p237-245, Elsevier Science Publishers B.V., Amsterdam.
- Ogden Environmental and Energy Services Company, Inc., 2000, Report on the Development of a Media Specific Soil Standard for Asbestos, prepared for Pennsylvania Department of Environment Protection
- USEPA, May 1992, Supplemental Guidance to RAGS: Calculating the Concentration Term, Publication 9285.7-081
- USEPA, Feb 2006a, Data Quality Assessment: Statistical Methods for Practitioners, Publication EPA QA/G-9S
- USEPA, Mar 2006b, On the Computation of a 95% Upper Confidence Limit of the Unknown Population Mean Based Upon Data Sets with Below Detection Limit Observations, EPA/600/R-06/022
- USEPA, Sept 2013, ProUCL manual – ProUCL Version 5.0.00 Technical Guide, EPA/600/R-07/041

Appendix A: Glossary

The glossary below provides meanings in simple terms to provide the reader with a broad understanding. It should not be seen as definitive.

- Asbestos Containing Materials (ACM): Materials that contain asbestos (e.g. cement boarding containing asbestos).
- Asbestos Related Work (ARW): Works that involve asbestos in soil at concentrations that are unlikely to create concentrations of asbestos in air that exceed 0.01 fibres/mL, but some precautions are required to protect workers (See BRANZ, 2017 for full details).
- Average: Expression of the central tendency of a dataset. Commonly synonymous with mean, but can refer to median or mode as well.
- Bias-Corrected Accelerated bootstrap (BCA): Bootstrap randomly resamples the dataset many times and calculates the mean of each resampling. From this a histogram of results from the resampled data are produced and from that an estimate of the variation of the mean. Bias-Corrected Accelerated bootstrap corrects for bias and skewness in a distribution.
- Box Plot: A simple box plot portrays the 25% quartile (base of box), 75% quartile (top of box) and median (line through the box). The whiskers above and below the plot extend to 1.5 times the interquartile range (the difference between the 75% and 25% quartiles).
- Chebyshev inequality: For normally distributed data, the 95% UCL is equivalent to two standard deviations. Chebyshev's inequality provides an equivalent for data that is not normally distributed.
- Class A/B Work: Work that involves asbestos in soil at concentrations that could theoretically create concentrations of asbestos in air that exceed 0.01 fibres/mL.
- Coverage: The likelihood that the calculated UCL will include the true mean of a population.
- Detection limit: Minimum reporting level for field screening and laboratory methods (0.001%).
- Kaplan-Meier Estimation Method (KM): Method developed for medical survival analysis (i.e. how long patients survive) used to assess statistical parameters (e.g. mean) of data containing a high proportion of non-detects.

- Mean: Sum of values in a dataset, divided by the number of samples.
- Non-detect: No asbestos detected in sample.
- Non-parametric: the distribution of the dataset does not follow a standard form.
- Outlier: Values that are not representative of the underlying population (equivalent to hotspots).
- Parametric distribution: The distribution of the dataset follows a standard form, e.g. normal or lognormal.
- Power: The ability to reflect accurately the true mean of a population, power is increased by using parametric methods with large data sets.
- Q-Q plot: Plot of the quantiles of two datasets against each other to assess whether they are similar. Figure 3 plots the quantiles of the example data against quantiles of a normal distribution.
- Quartile: 25% of the measured values are below the 25% quartile. 75% are lower than the 75% quartile.
- Robust Order Statistics (ROS): Method used to substitute non-detects with values based on the distribution pattern of the dataset (e.g. normal distribution).
- Skewness: Measure of asymmetry of the distribution of a dataset. There are a number of measures of skewness; the standard deviation of the log transformed dataset has been used in this paper.
- Unlicensed Asbestos Work (UAW): Works that involve asbestos in soil at concentrations that are theoretically unlikely to create concentrations of asbestos in air that exceed 0.01 fibres/mL and only minimal precautions are required to protect workers (See BRANZ, 2017 for full details).
- Upper Confidence Limit (e.g. 95% UCL): In simple terms, any randomly selected sample from the true population has a 95% probability of being below the 95 UCL.

Appendix B: Gy's theorem

In the second half of the 20th Century, Pierre Gy pioneered understanding of the errors that can occur when sampling heterogeneous materials; principally mining assays, but expanding to non-mining applications. One of the main principles is that materials containing discrete "nuggets" require more sample volume to characterise than materials with the same mass of mineral finely disseminated (i.e. 0.01g of asbestos distributed evenly throughout a soil will need less volume of sample to characterise than if the same mass of asbestos is clustered in a single fragment of boarding).

The formula used to generate Figure 2 is:

$$Ms = \frac{\mathrm{fcgl}d^3}{S_{FSE}^2}$$

Where:

- Ms = Mass of sample required to characterise soil
- f = Particle shape factor (coefficient of cubicity) - taken as 0.5 for laboratory analysis, assuming a rough spherical cluster of asbestos particles and 0.1 for field screening, assuming flat fragments of cement boarding (Minkkinen, 1987)
- c = Mineralogical (or composition) factor – approximates to the density of the nuggets divided by the concentration in the soil (Minkkinen, 1987). Density of the nuggets is taken to be 3000kg/m^3 (OEESC, 2000) divided by five, on the assumption that that particle is derived from degraded particle board (i.e. a density of 600kg/m^3 assumed)
- g = Particle size distribution factor – 0.5 for laboratory analysis (particles passed through a 2mm sieve) and 0.25 for unsorted field screened particles (Minkkinen, 1987)
- l = Liberation factor = 1 for particles that are completely detached from the soil (Minkkinen, 1987)
- d = largest particle size (sieve mesh diameter that would retain 5% of the sample) – taken to be 2mm for laboratory analysis and 10mm for field screening
- s^2_{FSE} = Standard deviation of the sampling error – taken to be 30% from typical relative percentage difference for blind replicates (MfE, 2011)

Appendix C: Identifying outliers – example

An example dataset for asbestos in soil laboratory results from the same soil horizon is provided below.

Results (%)	0.031	0.029	0.023	0.006	0.003	0.003	0.003	0.002	0.002	0.0011
	0.0008[3]	0.00044[4]	0.0002[4]	0.0001[4]	0.0003[4]	ND	ND	ND	ND	ND

There is insufficient data for statistical analysis for outliers, leaving three other methods to assess whether outliers are present:

- Q-Q Plot
- Box Plot
- Interrogate the ground model

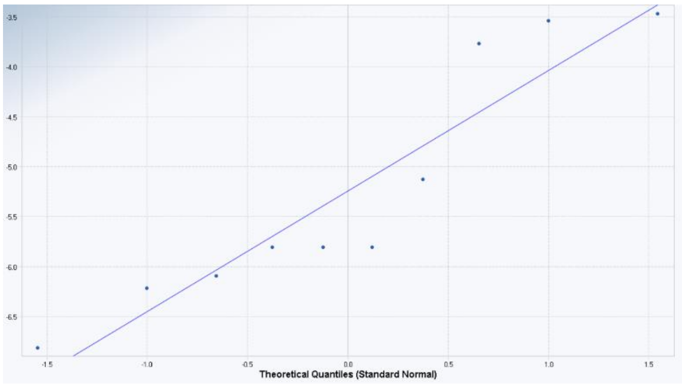

Figure 3: Q-Q plot for example data (screenshot from ProUCL software)

Q-Q Plot

All the points plot close to the theoretical lognormal trend, as shown in Figure 3. Therefore, examination of the Q-Q plot does not indicate the presence of outliers.

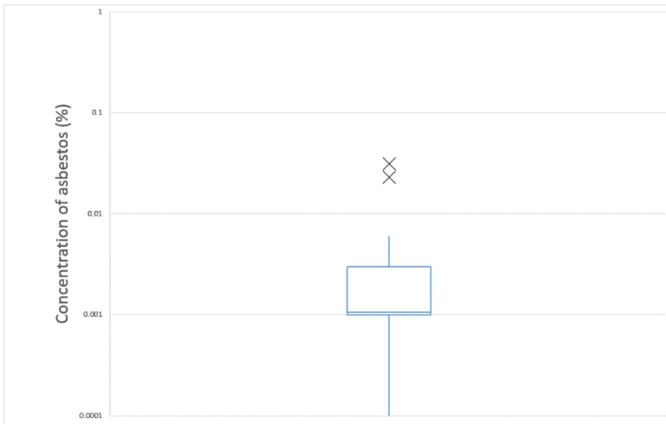

Figure 4: Box plot of example data (Note: Median is less than 0.001%)

Box Plot

A box plot indicates that the 0.031% and 0.029% values could potentially be outliers (Figure 4): values extending between 1.5 and 3 times the interquartile range (i.e. within the upper or lower whiskers) are typically considered "mild" outliers, while values greater or less than 3 times the interquartile range (i.e. beyond the whiskers) are considered "extreme" outliers (ITRC, 2013).

[3] Below Detection Limit of 0.001%

Ground model

We will assess whether outliers are present looking at two different ground models with exactly the same dataset (see Figure 5):

- Example 1 – open site with a 1m thick layer of imported fill.
- Example 2 – as Example 1, but with a former building in one corner, and a different spread of the data across the site.

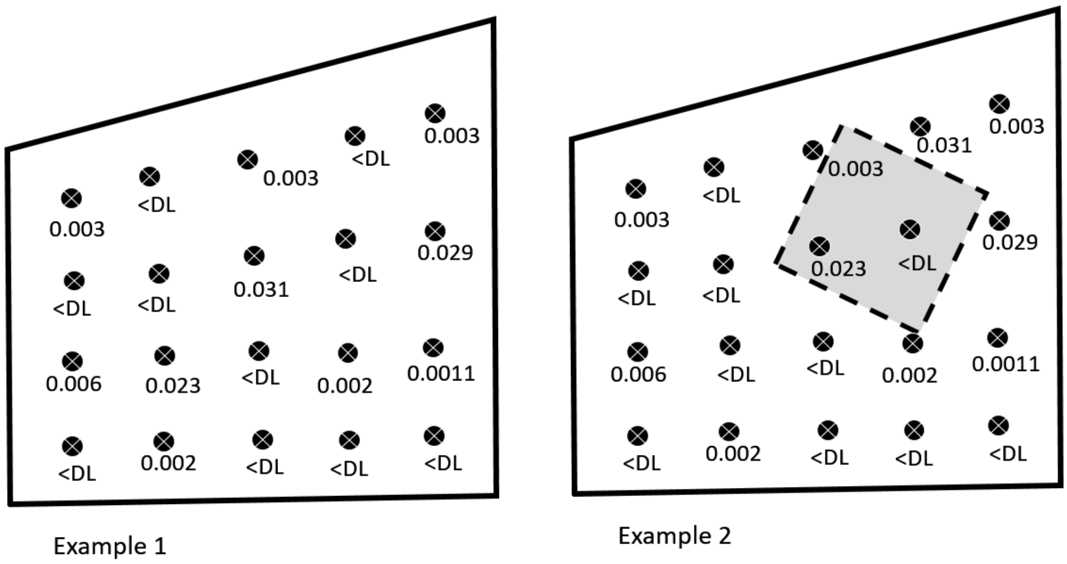

Figure 5: Plan view of example sites (asbestos results in %)

Both sites are to undergo commercial redevelopment that will seal the asbestos contaminated soil beneath concrete hardstanding and buildings.

In Example 1, the data is distributed across the site and there is no evidence from the ground model to suggest the presence of outliers. In Example 2, there is a cluster of high results around the former building, which from the ground model may represent a hotspot.

It is difficult to assess the boundary of the hotspot in Example 2. Conservatively for this example, the boundary of the hotspot could be placed so that there is one non-detect

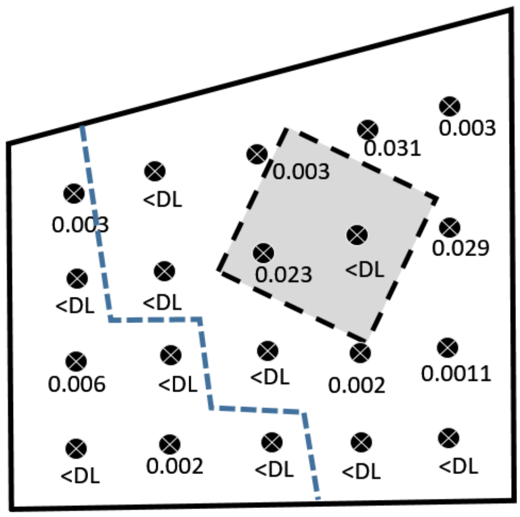

Figure 6: Dashed line denotes notional hotspot boundary

between each of the high results[4] and the edge of the hotspot (see Figure 6), although there are other options and further testing would likely be required to assess which is appropriate.

[4] Bearing in mind that we would expect at least every second sample to be a non-detect even in a high concentration area, based on at least 50% non-detects generally being reported for asbestos in soil sites.

Appendix D: Example UCL calculation

Calculation of a UCL is undertaken for Example 1 below. For Example 2 the distribution of the results is likely not independent (i.e. they increase near the building) and the area outside the hotspot has less than 10 samples. Therefore a UCL has not been calculated for this Example 2.

The dataset has 20 samples, is lognormal, with a skewness of 1.2[5] and 50% non-detects. From Table 1, the BCA is an appropriate method, which using ProUCL calculates a 95% UCL of 0.01. This data can therefore be used as a line of evidence that the soil disturbance works can be undertaken as Asbestos Related Works. The data is plotted in Figure 7 below, along with results for each of the UCL methods.

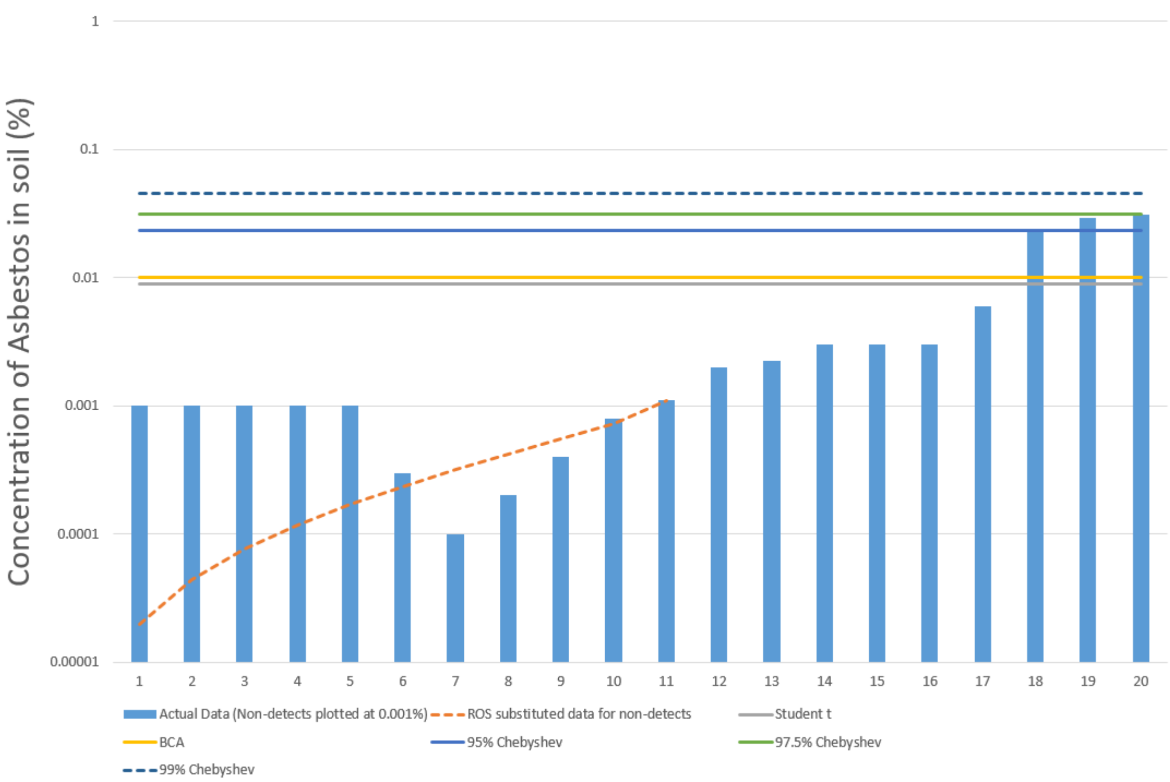

Figure 7: Dataset with ROS substitution data and calculated UCLs (note the five results on the left hand side have been plotted at the detection limit of 0.001% and the ROS substitution has only been used on these five results to calculate the UCLs)

[5] Skewness taken as the standard deviation of the log transformed data