A Statistical Review of a Report on Lead Concentrations in Residential Soil near a Potential Source of Lead Contamination in Sacramento County

This review was prepared for the County of Sacramento Environmental Management Department (EMD) by Nayak Polissar, PhD and Moni Neradilek, MS

The Mountain-Whisper-Light Statistics, Seattle, WA 98112

June 29, 2018

Table of Contents

Preface

The following 10 points express our opinions based on a review of a December, 2017 report by Stantec Consulting Services, Inc.: Mangan Park Neighborhood Background Lead Study/James Mangan Rifle and Pistol Range. Prepared for: City of Sacramento Parks and Recreation Department (hereafter, “Stantec Background report,” “Stantec Report”, “Stantec, 2017”, or “Stantec Consulting Services, Inc., 2017”). Our opinions may change if additional information becomes available. While all of our points should be considered, double asterisks (“**”) mark the more important points. This review assumes that the readers are familiar with the noted Stantec report.

The information and analysis presented in this report have been prepared under the supervision of, and have been reviewed by, Nayak Polissar, PhD, Principal Statistician of The Mountain-Whisper-Light Statistics.

Nayak Polissar

Signature

Executive summary

(1) The work by Stantec described in their report (Stantec Consulting Services, Inc., 2017) includes differences in field procedures and, potentially, laboratory procedures between the Tier 1 and Background studies. The differences introduce substantial uncertainty into the data collected and, thus, should be taken into account when interpreting any results based on the data—including Stantec's results and the new results presented in this review report.

(2) The method of statistical analysis by Stantec is not suited to the data. Their methods do not address the uncertainty (variance) of the lead concentration measurements and the difference in variability of lead measurements between the Tier 1 and Background areas. The methods also do not provide an estimate of the uncertainty (margin of error) for the difference in lead concentrations between the two areas. Their assessment that the mean lead concentrations in the two areas are "statistically indistinguishable" (a term not defined in statistical practice) is not supported by their analysis.

(3) Using an appropriate method of statistical analysis and accepting the soil sampling and laboratory analyses as if they are free of various problems (or potential problems), the estimated difference between Tier 1 and Background mean soil lead concentration is 4 mg/kg (Tier 1 minus Background). Based on an appropriate calculation of uncertainty, the difference is unlikely to be more than 18 mg/kg. The estimated difference in medians is 6 mg/kg, and the difference is unlikely to exceed 20 mg/kg.

Our review is organized under 10 different topics ("Points"). Double asterisks ("**") mark the more important statistical points. Additional material on some of the points is included in the Appendix.

Stantec's conclusions

** Point 1: Stantec's conclusion of "statistically indistinguishable" differences.

Stantec's current methods do not support the conclusion that the average soil lead concentrations in the Tier 1 area and the Background area are "**statistically indistinguishable."**[1] The term "statistically indistinguishable" is not a scientific term nor an otherwise established term. Because of this undefined term we can only surmise what the authors of the Stantec report meant by it. Our notion (and the overall impression from the report) is that Stantec concludes that the lead concentrations in the Tier 1 and Background areas are, using defined statistical terms, either *identical*[2] or *equivalent*[3]. Such a conclusion (if that is what Stantec means by "statistically indistinguishable") is not supported by Stantec's statistical comparison of the soil samples. Stantec's statistical analysis does not assess equivalence of lead concentrations in the Tier 1 and Background areas.

A statistically supportable analysis of the difference in lead concentration between the two areas is offered in Point 7, below. Based on the method of analysis in Point 7, the confidence interval (margin of error) around the estimated difference (Tier 1 minus Background mean lead concentration) can be used to determine if the true (unobservable) difference in mean soil lead concentration between the Tier 1 and Background areas may be of concern. In order to establish that there is actually statistical equivalence in lead concentration between the two areas, a maximum acceptable true difference between the mean concentrations of the two areas needs to be selected. This would be a difference that—based on the impact on health or other considerations—can be accepted as not large enough to violate the notion of equivalence between average lead concentrations in the Tier 1 and Background areas (see Point 7)[4]. Finally, the results of the new analysis in this review report have to be viewed in light of data collection limitations and modeling assumptions[5].

---

[1] "Statistically indistinguishable" is first used in the "Executive Summary" on page 1.1 of Stantec, 2017 and appears elsewhere in the document, including in the "Conclusions" on page 7.14.

[2] "Identical" means that some metric, such as the mean or the median, is exactly the same in both areas. The identity claim is impossible to prove with data that have measurement error. In the rest of our review we therefore focus only on assessing equivalence, which is both a meaningful and a feasible objective in the context of this study.

[3] "Equivalent" means that a difference in a metric (such as the mean or the median) is very likely to fall within a pre-specified ecologically meaningful margin, typically small.

[4] In medicine (where equivalence trials are common) two treatments can be designated as equivalent after a clinical trial if the difference in the average medical outcome (e.g., percent of patients with a positive outcome) is very unlikely to be more than a pre-specified amount. E.g., based on the data, a true difference between treatments of more than 5% in the number of patients with a positive outcome is very unlikely.

[5] The analysis in Point 7 takes the data at face value, and specific distributional assumptions are assumed to be true or, at least, not severely violated. See Point 10 for discussion about and limitations of the analysis in Point 7.

Data collection

Point 2: Property selection and representativeness.

The Background area contains 54 properties and the Tier 1 area contains 41 properties. We could not determine if the selected properties (28 properties from Background—52%—and 31 from Tier 1—76%) are representative of all properties in their respective areas. Permission to carry out soil sampling was obtained for only about half of the properties in the Background area. In any survey, a participation rate this low should be accompanied by some consideration of possible selection bias. Do the properties with a sampling agreement represent—in an unbiased way—all of the properties in terms of lead concentrations? That question was not addressed in the Stantec report.

Secondly, field procedures differed between the Tier 1 and Background surveys. We understand that the content of communications with the property owners to obtain permission for soil sampling was quite different for the Tier 1 and the Background studies. And, the extent of media coverage of the lead issue was quite different at the time of each soil survey. The letter sent to residents of the Tier 1 area by the City of Sacramento Department of Parks and Recreation, requesting access for soil sampling, raises the possibility that lead levels in the residents' yards is above a level considered acceptable by the State of California. (See the Parks and Recreation Department letter in the Appendix, Figure A0.) On April 20, 2017, the Sacramento Bee newspaper carried the headline "*Residents near park slam Sacramento officials about lead contamination.*"[6] A video filmed at a community meeting about the lead shows a resident upset with the Sacramento officials. (To view the video use the link in the preceding footnote). We understand that the public atmosphere at the time of requests for Background soil sampling was relatively calm compared to the atmosphere surrounding the Tier 1 sampling[7]. The difference in participation rates of 76% for Tier 1 and 54% for Background properties may be related to the different climate at the two survey periods, and the difference is statistically significant (i.e., unlikely to be due to chance)[8]. This difference in survey participation rates raises the possibility that—comparing Tier 1 and Background—the residents who agreed to soil sampling (and the cumulative history of activities on their soil) may differ between Tier 1 and Background in a way that is related to the soil lead concentrations on their properties.

Representativeness is an important assumption in the Stantec analysis, and our numeric comparisons in Table 1 also assume, essentially, that the participating homes were the result of a random selection process. Assessing representativeness may be difficult, but the issue was not considered in the Stantec Background report—either in an analysis or as a discussion point.

---

[6] Accessed online May 22, 2018 at http://www.sacbee.com/news/local/news-columns-blogs/city-beat/article73003277.html. Also see coverage at https://capitalandmain.com/a-neighborhood-gun-ranges-legacy-lead-contamination-1208 (accessed on May 22, 2018).

[7] This statement is based on discussions with Sacramento County Environmental Management Department staff.

[8] Comparison of 76% vs. 52%: $p = 0.02$, Fishers Exact Test. $P < 0.05$ is commonly considered "statistically significant."

** Point 3: Coring vs. trowel.

There may be a relative bias in measured lead concentrations between soil samples collected by coring (Tier 1) vs. samples collected by trowel (Background). The combination of the instructions and the soil collection device used in the Tier 1 area (4 cm diameter coring device, sampling to a depth of 3 cm[9]) suggests a more reproducible sample geometry and volume than in the Background area (trowel sampling to a depth of 1 inch[10]). Note the difference in sampling depth between the Tier 1 and Background areas: 3 cm vs. 1 inch (=2.54 cm), respectively. The Tier 1 specified sampling depth is 18% greater than the Background sampling depth. If lead concentration decreases with depth, the Tier 1 samples (and their deeper extent) could be biased toward lower concentrations. However, and perhaps more important, it is uncertain how a trowel method for soil extraction could consistently duplicate the depth and geometric shape of a core sample.

The distinct differences in soil sampling methods between the two areas could lead to a bias in the estimated lead concentration means and other statistics for Tier 1 vs. Background. There could also be differences between the precision of the coring measurements and the precision of the trowel measurements. There appear to be differences in the within-property variability between the Background and Tier 1 areas. As shown in Figure 1, compared to the Tier 1 measurements the Background measurements are more variable among the replicates within a property. This difference could be attributed to the differences in precision of soil sampling using coring vs. troweling, other differences in data collection methods (see Points 4 and 5) or truly larger within-property variability for the Background properties, or there could be a combination of the reasons noted.

The differences in within-property variability between the Tier 1 and Background areas generate a difference in the distributions of lead concentration values. The larger variability of the Background area's within-property replicates widens the Background area's between-property distribution. The difference in variability between the two areas needs to be addressed in the analysis. Our new analysis addresses the differing within-property variability analytically—an adjustment not carried out by Stantec.

[9] Tier 1 soil sampling instructions: "Sample increments will be collected from the upper 3 cm of soil using a nominal 4-cm-diameter sampling barrel." Quoted from a letter to Karl Kurka, Environmental Program Manager, City of Sacramento, from Neil Doran, Stantec, September 21, 2016; digital file title: "Directive 4c Workplan 09-21-16."

[10] "At each sampling station, three replicate samples were collected from shallow surface soils (upper 1 inch of soil). Samples were collected using steel trowels." From Stantec, 2017, page 3.4.

** Point 4: Differences in the incremental sampling sample sizes between Tier 1 and Background.

Each Tier 1 lead concentration replicate measurement for a given property was based on compositing 30 soil samples from the property, while each Background lead concentration measurement was based on a composite of only 5 samples. Because of the smaller number of subareas sampled within the Background properties than within the Tier 1 properties, the Background sample replicates (prior to compositing) are also expected to have more diverse lead concentrations, on the average, than in the Tier 1 area. The smaller number of samples per replicate in the Background area means that, compared to the Tier 1 area, there will be a larger average distance between the sub-area soil locations. This larger average distance, along with the smaller number of samples will lead to a larger expected measurement error and larger within-property variability for the Background area than for the Tier 1 area. The analysis in Point 7 verifies this larger variability in the Background area and uses a statistical method which accommodates the difference in variability.

Point 5: Laboratory effect: different laboratories were used for determining soil lead concentrations in the Tier 1 area and in the Background area.

Curtis & Tompkins, Ltd., Berkeley, California carried out the determination of lead concentrations for the Tier 1 area, and ESC Lab Sciences in Mount Juliet, Tennessee carried out the determinations for the Background area. Given the use of two different laboratories for the lead analysis, we would have liked to see more evidence in the Stantec report narrative that identical laboratory procedures were followed in the two laboratories. Particularly important to compare between laboratories would be a) the method of homogenizing soil samples, b) the mass of aliquots drawn for lead determination, and, c) the procedures for laboratory quality control and calibration to external lead standards[11]. If soil composites were not homogenized[12] as thoroughly in one laboratory as in the other (allowing a greater chance in one laboratory of selecting a lead "nugget" into the aliquot taken from the composite), or if a markedly greater mass of composite sample was chemically analyzed in one laboratory than in the other, then the variability of lead concentrations among the replicate samples would very likely differ between the two laboratories.

---

[11] See Martin, Rooney, Woods, & Woods, 2018, section 2.4 for an example from a peer-reviewed journal of documentation of laboratory procedures for determination of lead concentrations in blood. While samples of blood and soil would have different forms of preparation and would differ in other procedures, the cited article is offered to show a good level of detail in the narrative text describing laboratory methods. The following quote is from the article by Rooney, J et al: "Repeat analyses of control samples and other procedures were performed weekly to ensure high consistency and accuracy of blood lead concentration assessments."

[12] The Stantec materials available to us offer the following description of homogenizing. "The material passing through the #60 sieve is homogenized and distributed across a pan to a depth of 0.25- to 0.5-inch." (Extract from a letter to Karl Kurka, Environmental Program Manager, City of Sacramento, from Neil Doran, Stantec, dated September 21, 2016; digital file title: "Directive 4c Workplan 09-21-16.")

We raise the issue of laboratory procedures because some consideration of the lead concentration data obtained by Stantec suggests that composites were not completely homogenized prior to drawing the aliquot for lead determination—at least in one laboratory. The evidence for non-homogeneity is that there were two very different lead concentration values from two distinct samples from the same replicate soil composite (see, also, Point 9). An aliquot from one replicate composite from Background property BG-8 yielded a concentration of 1,370 mg/kg, and a second aliquot from the same composite yielded 210 mg/kg—more than a six-fold difference[13]. The main report of an important study such as this would be more complete by considering the potential laboratory effect on results. Given the importance of this study to the community, topics that should be addressed are the use of two different laboratories for analyzing soil from the Tier 1 and Background areas, respectively, and the hint of incomplete homogenizing of composite samples.

Data analysis

** Point 6: The UCL (upper confidence limit) is an inappropriate measure of lead concentration.

Stantec's use of the UCL as a measure of a property's lead concentration and the comparison of the Tier 1 and Background soil lead concentrations using UCLs is not meaningful or interpretable.

The Stantec report extensively uses the 95% upper confidence limit (UCL) of the mean as a metric to compare the Tier 1 and Background areas. A per-property 95% UCL (e.g., Plot 3 of Stantec, 2017) and a per-area UCL (e.g., Stantec's Table 1) does not estimate any relevant physical parameters about a property or an area. Both types of UCLs are affected by parameters that are not related to soil lead concentration. For example, the parameter of sample size is used in the calculation of the UCL, but the sample size is not a physical parameter about the property. These parameters should not be allowed to influence the estimate of the overall lead burden. Considering the Stantec use of UCLs, it would be possible for a property "X" to have a <u>higher</u> lead UCL than property "Y," while the means would be just the reverse—the mean of property "X" could be <u>lower</u> than the mean of property "Y." This kind of inconsistency does, in fact, happen in the Stantec data.[14] Thus, the UCLs are not an appropriate method of comparing lead concentration among individual properties or between the Tier 1 and Background sites.

---

[13] The two other replicate composites from the same property each had two aliquots drawn for lead determination, and the changes in lead concentration from the first to the second aliquot were relatively much smaller (1st and 2nd aliquot, per replicate: a) 189 mg/kg initially, followed by 150 mg/kg; b) 372 mg/kg initially, followed by 160 mg/kg).

[14] An example of inconsistent lead concentration results when comparing means vs. comparing UCLs can be found for properties BG-8 and BG-16 (using data from Table 1 of Stantec, 2017). BG-8 has a <u>*higher*</u> mean lead concentration than BG-16 (173 mg/kg vs. 145 mg/kg), whereas—just the reverse—BG-8 has a <u>*lower*</u> UCL than BG-16 (254 mg/kg vs. 276 mg/kg).

Instead of using UCLs as the metric, the analysis should only use readily interpretable and meaningful metrics, such as the arithmetic mean[15] or median of the between-property distribution of lead concentrations in the two areas. And, as noted in Point 7, the analysis should also account for measurement error.

** Point 7: Flawed inferential methods and some new calculations.

A major issue with Stantec's statistical analysis is the lack of proper statistical methods—methods that can be used to correctly infer the magnitude of the difference in lead concentrations between the Tier 1 and Background properties[16]. One methodologic aspect is that it is critical to have a confidence interval (margin of error) for the difference of means (or medians) between the two areas. The confidence interval is needed in order to assess whether or not there is equivalence of lead concentrations between the areas [17]. The confidence interval for an equivalence assessment is absent from the Stantec report. A commonly used statistical method (random effects models (Bates & Pinheiro, 2000), used in our analysis, below) can provide the appropriate confidence interval.

A key part of the appropriate methodology addresses measurement error—the common phenomenon that measured quantities (such as lead concentrations) include some "noise" (i.e., error) in the measured value. Even if a laboratory measures the lead concentration very precisely for a given soil sample, that measured lead concentration value for that sample inevitably will differ to some degree from the average soil lead concentration of the property. That is measurement error. There is measurement error embodied in each measured lead concentration value. The random effects model used in our analysis addresses the impact of measurement error and leads to a confidence interval that appropriately represents the uncertainty of estimates[18].

Within the random effects methodology used here, there is also an assumption that certain quantities have a normal ("bell-shaped") distribution. In order to evaluate the normal distribution assumption, we will consider random effects models that use lead concentrations

---

[15] We re-calculated and reproduced Stantec's estimates of the (arithmetic) mean lead concentration of 71 mg/kg and 66 mg/kg for the Background and the Tier 1 areas, respectively. Both are valid point estimates of the mean lead concentration in each area.

[16] For the presentation in this section we will assume that the soil lead measurements have been collected, physically handled and chemically analyzed in a way that does not introduce any bias into the difference in lead concentrations between Tier 1 and Background areas. This strong assumption implies that the soil collection methods (soil cores and soil troweling in Tier 1 and Background area, respectively) are completely interchangeable and, as methodologies, give equivalent results. If, on the other hand, there are systematic differences in lead concentrations between the two collection methods (coring device vs. trowel) and, possibly, different methods of composite preparation for analysis by the two laboratories, then the possibility of bias should be borne in mind when interpreting results. Non-random selection of properties for soil sampling is also a potential source of bias.

[17] Confidence intervals are needed to appropriately assess equivalence. In contrast, p-values are not a good indicator of equivalence. See the last paragraph of the Appendix material on Point 7 for a discussion of p-values.

[18] See the additional discussion of the random effects model in the Appendix under Point 7.

on the original scale (mg/kg) and, alternatively, on the logarithmic (log_{10}) scale. We will assess whether the original or the logarithmic scale provides a more valid confidence interval. We will also present a sensitivity analysis to evaluate the impact of different choices for handling Background property BG-8, which Stantec designated as having an outlier lead concentration. Stantec replaced the three initial lead concentration values from BG-8 with new values obtained from three new aliquots from the same three replicate soil composites.

An additional statistical consideration related to measurement error is that the measurement error may also vary across the properties[19]. The measurement error can be expressed empirically as a variance or standard deviation (SD) of the lead concentrations within a property. We found that the soil concentration data violated the assumption of equal variances among properties on the original (mg/kg) scale. (See the summary trend lines in Figure 1). Specifically, there were larger measurement errors for properties with higher lead concentrations. Additionally, the Background properties had a higher within-property variation in lead concentrations than the Tier 1 properties. The average within-property SD was very small (just above zero mg/kg) for Tier 1 properties with means less than 50 mg/kg but 2- to 3-fold larger for Background properties in the same range of less than 50 mg/kg. The within-property SD then increased for larger means (in both areas) with a higher mean SD for Background properties than for Tier 1 properties until about the 85 mg/kg mean. The curves to the right of mean 85 mg/kg are relatively poorly determined due to small sample sizes. However, for both Background and Tier 1 properties with mean >85 mg/kg the within-property SDs were relatively large (mostly >20 mg/kg) with substantially more of the properties with the largest measurement errors originating from the Background area.

[19] Heteroscedasticity is the term used to refer to unequal variance for different groups of measurements. Here, as will be noted from Figure 1, the variance of the three measurements within a property does vary among the properties. I.e., there is heteroscedasticity.

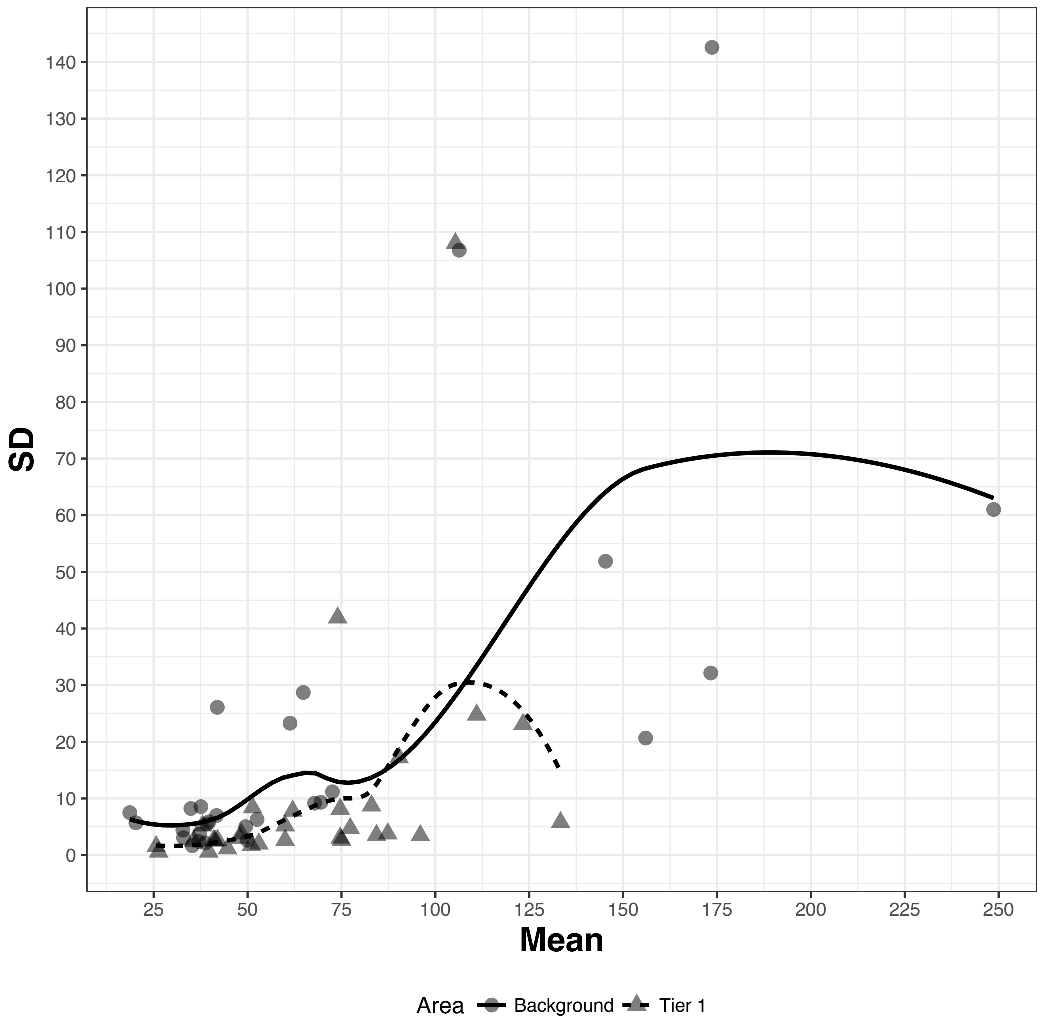

Figure 1. Within-property mean lead concentration (horizontal axis) vs. within-property standard deviation (SD, vertical axis) of the three replicate measurements per property. Background = dots and full line; Tier 1 = triangles and dashed line. Lines = trend in mean SD (estimated by the LOESS smoother, (Cleveland, Grosse, & Shyu, 1992).

Figure 1 illustrates that the within-property variation is larger for properties with high lead concentrations and that, in general, the Background area properties have larger within-property variation than the Tier 1 area properties.

The differences in the measurement errors between the Tier 1 and the Background properties are also illustrated in results from the random effects model. On the original scale the within-property SD for the Background area was 39 mg/kg, almost double the corresponding SD of 22 mg/kg for the Tier 1 area (numeric values are from Table A1 in the Appendix). A considerable difference in the SD was also found on the $\log_{10}$ scale (SD = 0.15 log mg/kg for Background vs. SD = 0.10 log mg/kg for Tier 1). The random effects model—as used here—assumes that the true SD is constant across the properties. The different SDs for the two areas were accommodated by fitting separate random effects models for data from the Tier 1 and the Background areas rather than combining the data in a single analysis.

In order to address the larger measurement errors for properties with higher lead concentrations (on the original scale) we considered the random effects model for data transformed to the $\log_{10}$ scale. It was apparent from that analysis, presented below, that the confidence intervals for the difference between Background and Tier 1 areas would be more valid using the $\log_{10}$ scale models than using the original scale models. See Appendix Figure A2 and the corresponding Appendix text for justification of the analysis using the $\log_{10}$ data.

The formulation of the random effects models (and, consequently, the validity of the confidence intervals) also relies on the normal (bell-shaped) distribution being a good approximation to certain distributions occurring in the analysis. The normal distribution assumption is substantially violated by the data on the mg/kg scale (the data as used by Stantec) but the assumption is much more acceptable on the $\log_{10}$ scale. (See Appendix Figure A3 and the corresponding Appendix text for details).

The random effects model estimates of mean and median lead concentrations for each area (and the derived difference of means or medians between areas) are presented in Table 1. The table shows the estimated means for the Background area (B), the Tier 1 area (T1) and the difference (T1-B). Confidence intervals and p-values were estimated using the parametric bootstrap (Davison & Hinkley, 1997). The calculated means on the original scale were the same as Stantec's estimates of the arithmetic mean lead concentrations of 66 mg/kg and 71 mg/kg for the Tier 1 and Background areas, respectively. Both values are valid estimates of the mean lead concentrations for the two areas, respectively, but (as pointed out above) the confidence intervals on the original scale have limited validity. The estimated difference of the arithmetic means (T1-B) was -4 mg/kg[20].

Estimated values from the model on the $\log_{10}$ scale are more appropriate. The model estimates a mean of 1.73 $\log_{10}$ mg/kg for the Background area and a mean of 1.78 $\log_{10}$ mg/kg for the Tier 1 area. The difference of means (T1-B) was 0.05 $\log_{10}$ mg/kg (95% CI: -0.07 to 0.16 $\log_{10}$ mg/kg). If we assume (plausibly) that the between-property distribution of the $\log_{10}$ transformed lead concentrations is normally distributed in each area, we can transform the means back to the original scale (see Appendix material on Point 7 for details). From this transformation we estimated a mean of 59 mg/kg (95% CI: 45-73 mg/kg) for the Background area and a mean of 62 mg/kg (95% CI: 53-72 mg/kg) for the Tier 1 area. The estimated difference of area means (T1-B) was 4 mg/kg[21] (95% CI: -13 to 18 mg/kg).[22]

---

[20] The difference of -4 mg/kg from the subtraction of the point estimates of 66 and 71 mg/kg is due to rounding.
[21] Note that the T1-B difference of the arithmetic means estimated by the model on the original scale was -4 mg/kg (See Table 1). The discrepancy between the two models (+4 mg/kg for the $\log_{10}$ model vs. -4 mg/kg for the original-scale model) does not imply an inconsistency between the two models, because the confidence intervals around the point estimates are relatively wide (a width of at least 31 mg/kg). Identical results for the two models would be guaranteed if the distributions were exactly log-normal and if the sample size was infinite. The noted difference in estimates between the two models (-4 mg/kg vs. +4 mg/kg) for these data could, therefore, be attributed either to departures from the log-normal distributions or to the small sample size (or to both).
[22] All statistical calculations in this report were carried out using the R software (R Core Team, 2018) version 3.4.0. Random effects models were implemented using the R package 1me4 (Bates, Bolker, Maechler, & Walker, 2015).

Do the Tier 1 and Background areas have an equivalent mean lead concentration? That question can be addressed as follows. Two means can be considered statistically “equivalent” in accepted practice if the difference between the means is less than a pre-specified bound. This bound is typically small and would not be considered of practical importance. Therefore, we would conclude an equivalence of the T1 and Background means, if a difference in those means of up to 18 mg/kg (the upper side of the 95% confidence interval) would be considered not of practical importance.

The $\log_{10}$ model can be also used to estimate the median on the original scale. When the means on the $\log_{10}$ scale are exponentiated, geometric means are obtained. If we further assume that the distributions on the $\log_{10}$ scale are symmetric, the geometric means coincide with the medians on the original scale. Under this symmetry assumption (which is reasonable as an approximation), we estimated a median of 54 mg/kg (95% CI 42-67 mg/kg) for the Background area and a median of 60 mg/kg (95% CI 52-70 mg/kg) for the Tier 1 area. The estimated difference of the two area medians (T1-B) was 6 mg/kg (95% CI: -10 to 20 mg/kg). Similar to our discussion on equivalence of means, above, we would conclude an equivalence of the true Tier 1 and Background medians, if a difference in those medians of up 20 mg/kg would be considered not of practical importance.

Neither the difference between the Tier 1 and Background means nor the difference between their medians is statistically significant (p = 0.6 to 0.7 in Table 1). This statistical non-significance, considered alone, does not show that the true (unobserved) mean soil lead concentrations are equivalent between the two areas. The confidence intervals need to be considered (as we have done, above) to determine whether the maximum likely difference in the mean soil lead concentration between the two areas is of no concern—i.e., “equivalent” in a practical sense. The p-values in Table 1 are based on a better method for calculating the level of statistical significance than in Stantec’s report. Stantec’s calculated p-values were based on the Mann-Whitney test, which has strong assumptions that are unlikely to apply to these data. See the Appendix, Point 7, for additional comments on Stantec’s use of the Mann-Whitney test and on the interpretation of p-values.

In summary, it appears very unlikely that the either the mean or median lead concentration in the Tier 1 area exceeds the corresponding value in the Background area by more than 20 mg/kg. However, see Point 10 for limitations of these analyses.

Table 1. Estimated mean and median lead concentrations for the Background and Tier 1 areas based on random effects models. Levels of statistical significance (P-values, P) are noted.

	Estimate (95% CI)	*P*
Arithmetic mean based on analysis using data on the original scale (mg/kg)*		
Background (B)	71 (51-91*)	
Tier 1 (T1)	66 (56-76*)	
Difference (T1-B)	-4 (-28 to 17*)	0.7*
Arithmetic mean based on analysis using data on the $\log_{10}$ scale		
Background (B)	1.73 (1.62-1.83)	
Tier 1 (T1)	1.78 (1.71-1.84)	
Difference (T1-B)	0.05 (-0.07 to 0.16)	0.6
Arithmetic mean expressed on the original scale (mg/kg) based on analysis using data on the $\log_{10}$ scale.		
Background (B)	58 (45-73)	
Tier 1 (T1)	62 (53-72)	
Difference (T1-B)	4 (-13 to 18)	0.7
Median expressed on the original scale (mg/kg) based on analysis using data on the $\log_{10}$ scale.**		
Background (B)	54 (42-67)	
Tier 1 (T1)	60 (52-70)	
Difference (T1-B)	6 (-10 to 20)	0.6

Notes: Confidence intervals and p-values were calculated using the parametric bootstrap with 1,000 replicates. These analyses assume normality of the within-property and between-property distributions (on the respective scales). That assumption appears to be violated by the analysis for the confidence intervals and p-value for the arithmetic mean based on data on the original scale (mg/kg).

**As described elsewhere, the within-property and between-property distributions substantially deviate from the normal distribution and, thus, the confidence intervals and p-values for the analysis of data on the original scale are possibly invalid.*

*** The estimated median assumes a symmetric distribution on the log scale. Under this assumption, the median is also an estimate of the geometric mean.*

Table 2 shows a sensitivity analysis in regard to handling of the BG-8 property, which Stantec cites as having an outlier lead concentration value. (See Point 8, below, for a discussion of the "outlier.") Our primary analysis used the replacement values for property BG-8 (150, 160 and 210 mg/kg). These three new values are based on resampling from the original three composite samples. Our first sensitivity analysis uses the original values for property BG-8 (189, 372 and 1,370 mg/kg for the three replicates, respectively) and our second sensitivity analysis used all values (original and new, a total of six values). The results in Table 2 show that the different ways of handling the BG-8 property lead concentrations had a relatively minor impact on the mean and median estimates, relative to the uncertainty in the parameter estimates (i.e., the width of the confidence intervals). The three columns of analyses in Table 2 show a range of differences of means (T1 – B) from 0 to 4 mg/kg on the original scale and a range from 4 to 6 mg/kg for medians. These ranges are very tight compared to the margin of error (width of confidence intervals).

Table 2. Sensitivity analysis to examine alternative handling of the BG-8 property lead concentrations. Estimated means, differences of means and confidence intervals (CI) for the Background and Tier 1 areas using random effects models on the $\log_{10}$ scale with back-calculation of values to the original scale (mg/kg).

	*Primary analysis (**new BG-8 values**)*	*Sensitivity analysis #1 (**original BG-8 values**)*	*Sensitivity analysis #2 (**all BG-8 values**)*
	Estimate (95% CI)	*Estimate (95% CI)*	*Estimate (95% CI)*
Arithmetic mean on the original scale			
Background (B)	58 (45-73)	62 (47-81)	60 (47-78)
Tier 1 (T1)	62 (53-72)	62 (53-72)	62 (53-73)
Difference (T1-B)	4 (-13 to 18)	0 (-21 to 18)	2 (-18 to 18)
Median on the original scale (also estimates the geometric mean)			
Background (B)	54 (42-67)	56 (43-73)	55 (43-71)
Tier 1 (T1)	60 (52-70)	60 (52-70)	60 (51-70)
Difference (T1-B)	6 (-10 to 20)	4 (-14 to 21)	5 (-14 to 20)

Notes: These analyses assume normality of the within-property and between-property distributions (on the $\log_{10}$ scale). The estimated median assumes a symmetric distribution on the log scale. Under this assumption, the median is also an estimate of the geometric mean. Confidence intervals and p-values were calculated using the parametric bootstrap with 1,000 replicates. Both assume normality of the within-property and between-property distributions on the $\log_{10}$ scale.

Point 8: Spatial Correlation.

The Stantec analysis and our analysis assume no spatial correlation of lead concentration among properties. We examined this assumption. Figures A4 and A5 show the spatial distribution of the three-replicate mean concentrations in the Background and Tier 1 areas, respectively. Figure A6 shows the corresponding variograms, which are relations of distance between a pair of properties and the covariation of their mean lead concentrations (Cressie, 1993). There is weak, if any, evidence that the assumption of no spatial correlation was violated. We therefore felt safe to proceed in our analysis assuming no spatial correlation.

Point 9: Outlier handling.

Stantec's handling of Background property BG-8 as having an outlier does not have an adequate justification for the replacement of three lead concentration values. A relatively large lead concertation is not *prima facie* in error and to be excluded, and, as noted below, the method of outlier detection used by Stantec is not appropriate. There is no laboratory or acceptable technical rationale provided as to why three original BG-8 lead concentration values should be considered to be in error due to the presence of one large value. Also, given the acceptable values of two out of the three original replicate values for property BG-8, why were the two values discarded along with the putative outlier? The analysis could have been carried out using just the two remaining replicates, whose lead concentrations fall in a range that was very common among the remaining properties. Our analysis (Table 2, discussed in Point 7) showed that the handling of property BG-8 (using the new values, the original values or both) had limited impact on the means and medians estimated by the random effects model on the $\log_{10}$ scale. Thus, post hoc, it seems to have mattered little that Stantec changed the outlier value, but without the sensitivity analysis (Table 2) the user would not know that the change of data would matter little.

Stantec's use of the 1.5 interquartile range (IQR) outlier rule is ad-hoc and is not sufficiently supported by either a relevant reference or a thorough statistical justification. Our first critique of the 1.5 IQR method is that it is highly dependent on the shape of the true population distribution. Another fault of the IQR (which can be shown analytically) is that—under repeated sampling from a population—a certain fraction of all observations will inevitably be designated as outliers.

Limitations

** Point 10: Some disclaimers and limitations.

The results of our random effects analysis, although more appropriate for the data than Stantec's analysis, will still have to be viewed as having some non-statistical limitations, as follows. (These limitations would apply to any analysis using the Tier 1 and Background lead concentrations.)

Our analysis cannot address biases that may have been introduced to the comparison of Tier 1 and the Background lead concentrations from the following factors:

- There is a potential for selection bias in properties included in the study. Our understanding is that the public atmosphere, media coverage available to the public and communications to the residents requesting permission for soil sampling on properties differed between the periods of the Tier 1 and Background studies. The percentage inclusion of properties in the Background area (52%) was substantially lower than in the Tier 1 area (76%), with statistical significance. The residents' personal decisions about whether or not to participate in the soil sampling may have delivered a different kind of resident (with their yard and soil) to the study from the Tier 1 area compared to the Background area.
- There were different soil extraction methods for the Tier 1 area (soil cores) and Background area (trowel). The trowel is manifestly not as precise as the core.
- Different laboratories were used for processing the composited samples, leading to the possibility of differences in procedures that may have affected the measured lead concentrations from soil samples and may have introduced a bias.
- The Tier 1 and the Background areas differed in the number of locations composited for replicate samples at each property. Thus, the variation among replicates from a property would be (and was found to be) larger in the Background areas.

Our analysis also does not factor in lead data other than Stantec's sampling of the Background and Tier 1 areas. We have not reviewed and commented upon Stantec's analysis of and reports on lead in the Sacramento urban environment and the alleged impact of Sacramento airport, but that is likely to be a helpful component of their report. It seems likely that the environmental sources of lead described by Stantec (aside from the gun range) would have affected the Tier 1 and Background areas equally.

The estimated parameters and their confidence intervals presented in this review report rely on some distributional assumptions. As examined by the model diagnostics (Figures 2 and 3) these distributional assumptions are reasonable given the data. However, we cannot be 100% confident about the assumptions because the dataset is relatively small and the assumptions cannot be thoroughly checked. The assumptions play a role in the back-calculation of the means from the values on the $\log_{10}$ scale to the values expressed on the original scale (mg/kg). The normality assumptions in the random effects model are reasonable but not completely verifiable in this limited dataset. The back-calculated estimates and confidence intervals of the arithmetic means assume that the between-property distribution (on the log_{10} scale) can be well approximated by the normal distribution. The back-calculated estimates and confidence intervals of the medians rely on an assumption that the unobservable between-property distribution of lead concentrations (on the log_{10} scale) is symmetric. The assumption is plausible but not certain. Finally, the validity of the median estimates and the confidence intervals for the mean and median lead concentrations rely on the assumption that the true within-property variance of lead concentrations (on the log_{10} scale) does not vary among the Tier 1 properties and, as well, does not vary among the Background properties[23]. Finally, confidence intervals from the analysis on the original scale (mg/kg) are presented for completeness (Table 1) but their validity is in question because the homoscedasticity and normality assumptions are clearly violated, as shown in this review report.

Conclusions

The Stantec statistical analysis is not suited to the data, because it does not properly address the greater variability of the Background data (compared to Tier 1 data). The data analysis presented in this review report does address the differential variability. The numeric value of the estimates by Stantec and by The Mountain-Whisper-Light Statistics for the difference between average lead concentrations in the Tier 1 and Background properties are similar (given the margin of error). However, the estimates provided in this review report can be used more confidently, given the use of an analysis methodology that is consistent with the data. Further, the presentation of confidence intervals in this review report can be used to assess whether the maximum likely difference between Tier 1 and Background is of concern.

Substantial non-quantitative uncertainty is attached to the comparison of soil lead concentrations between Tier 1 and Background due to the different study procedures in the two areas (see Point 10). However, if the data are accepted as free from the problems noted, the estimated difference between Tier 1 and Background mean soil lead concentration is 4 mg/kg (Tier 1 minus Background), and the difference is unlikely to be more than 18 mg/kg. The estimated difference in medians is 6 mg/kg, and the difference is unlikely to exceed 20 mg/kg.

[23] If a collection of properties do not have the same true constant within-property variance of lead concentrations, it is referred to as "heteroscedasticity." The observed variances, of course, may differ from property to property, but, if there is no heteroscedasticity, then the expected variance is the same for every property. If the true variances are all equal, it is referred to as "homoscedasticity."

References cited

Bates, D. M., & Pinheiro, J. C. (2000). *Mixed-Effects Models in S and S-PLUS.* New York: Springer.

Bates, D., Bolker, B., Maechler, M., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using 1me4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01

Cleveland, W. S., Grosse, E., & Shyu, W. M. (1992). Local Regression Models. In T. J. Hastie, & J. M. Chambers (Eds.), *Statistical Models in S.* University of Michigan: Wadsworth & Brooks/Cole.

Cressie, N. (1993). *Statistics for Spatial Data.* University of Michigan: Wiley.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and their Application.* Cambridge: Cambridge University Press.

Martin, M. D., Rooney, J., Woods, J. S., & Woods, N. F. (2018). Genetic polymorphisms of GRIN2A and GRIN2B modify the neurobehavioral effects of low-level lead exposure in children. *Environmental Research*, in press.

R Core Team. (2018). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Retrieved from R Foundation for Statistical Computing: https://www.R-project.org

Stantec Consulting Services, Inc. (2017). *Mangan Park Neighborhood Background Lead Study/James Mangan Rifle and Pistol Range.* Sacramento: City of Sacramento Parks and Recreation Department.

Appendix

Additional material for selected points.

Point 1: No additional material.

Point 2:

City of
SACRAMENTO
Parks and Recreation

September 26, 2016

RE: Request for Permission to Test Soil in Your Yard for Lead Contamination

Dear Property Owner/Resident:

As you may know, the City of Sacramento previously operated the Rifle and Pistol Range located at James Mangan Park at 2140 34th Avenue. The range was closed in December of 2014. While the range was being operated, the firing of lead bullets inside the range created lead dust debris that was subsequently tracked outside by rarge users. The dust was also deposited on the outside of the building and the roof from the air exhaust system, and that dust washed down during the rain into the soil and the parking lot areas that surround the building. The lead dust is only hazardous if it is ingested.

The City has been working with the County of Sacramento's Environmental Management Department and the California Department of Toxic Substances Control since April 2016 to conduct testing on exterior building surfaces, hardscape areas and soil around the range and in the park to determine if the level of lead on those surfaces is unsafe. The testing has shown that lead contamination in the soil is limited to the park areas closest to the range building. Removal of the contaminated soil and hardscape in these areas is nearly complete.

Although lead is a heavy metal, it is possible that the lead dust from the exhaust system may have been carried by the wind and deposited beyond the range roof. The County has directed the City to test soil in the yards of homes located north of the range to determine if lead concentrations exceed the State's standard of 80 milligrams per kilogram (80mg/kg) for residential areas. Soil testing consists of collecting 30 small soil samples from landscaped areas using a hand-held soil boring tool, and sending the samples to a laboratory for analysis. Sample collection should not cause any damage to the landscaping. **The purpose of this letter is to request permission to test the soil on your property.**

Please note: lead is commonly found in the environment and lead paint was used in older homes, so a positive sample may not necessarily indicate contamination from the range. The City will notify a property owner if the lead testing indicates that the lead in the soil exceeds the State's standard and discuss soil remediation options with those affected. Complete testing results will be sent to the County and the information will be available to the public.

Parks and Recreation Department
916-808-5200
915 I Street, Third Floor
Sacramento, CA 95814

Figure A0. Letter requesting access for soil sampling on Tier 1 properties.

Points 3-6: No additional material.

Point 7:

Additional comments on random effects models and statistical issues associated with use of the random effects model

The random effects model used in this review report is needed to account for measurement error in the lead concentration data. The measurement error is reflected in the variation of the lead concentrations across the three replicate measurements within a property. Because of measurement error, none of the three replicate measurements will exactly equal the true property mean lead concentration. Additionally, the mean of the three observed replicate measurements will not exactly equal the true property mean lead concentration either. The significance of the measurement error is that the between-property distribution of the averages (means) of the three replicate measurements will be biased compared to the distribution of the true per-property means (an unobservable, true distribution). Specifically, the tails of the distribution of the three-replicate averages will be wider compared to the true between-property distribution. The random effects model can estimate the measurement error and adjust the between-property distribution accordingly to yield a correct between-property distribution of lead concentrations.

Table A1. Results of random effects models by study area (Background vs. Tier 1), presented for original and $\log_{10}$ scales.

	Original scale (mg/kg)			**Log_{10} scale**		
Area	*Arithmetic mean*	*Between-property SD*	*Within-property SD*	*Arithmetic mean*	*Between-property SD*	*Within-property SD*
Background	71	52	39	1.73	0.26	0.15
Tier 1	66	25	22	1.78	0.17	0.10

The model parameters estimated in the random effects models are the mean, the between-property SD and the within-property SD. These parameters for the two areas are listed in Table A1 (separately for the original and $\log_{10}$ scale models). The results show that the background properties were more variable both in terms of the between- and the within-property variation, which can be seen by comparing the SD values between the Background and Tier 1 areas in Table A1. The between-property SD (on the $\log_{10}$ scale) is utilized in back-transforming the means from the $\log_{10}$ scale to the original scale as follows: $mean_O = 10\text{^}(mean_L + SD_L^2/2)$, where $mean_O$ is the calculated mean on the original scale, $mean_L$ is the mean on the $\log_{10}$ scale and SD_L is the SD on the $\log_{10}$ scale. In the parametric bootstrap both $mean_L$ and SD_L are recalculated for each bootstrap realization.

Figure A1 shows that the Tier 1 and the Background distributions have different shapes (the background distribution is wider) which explains the differences between the means derived from the original scale (mg/kg) vs. derived from the log scale (presented in Table 1).

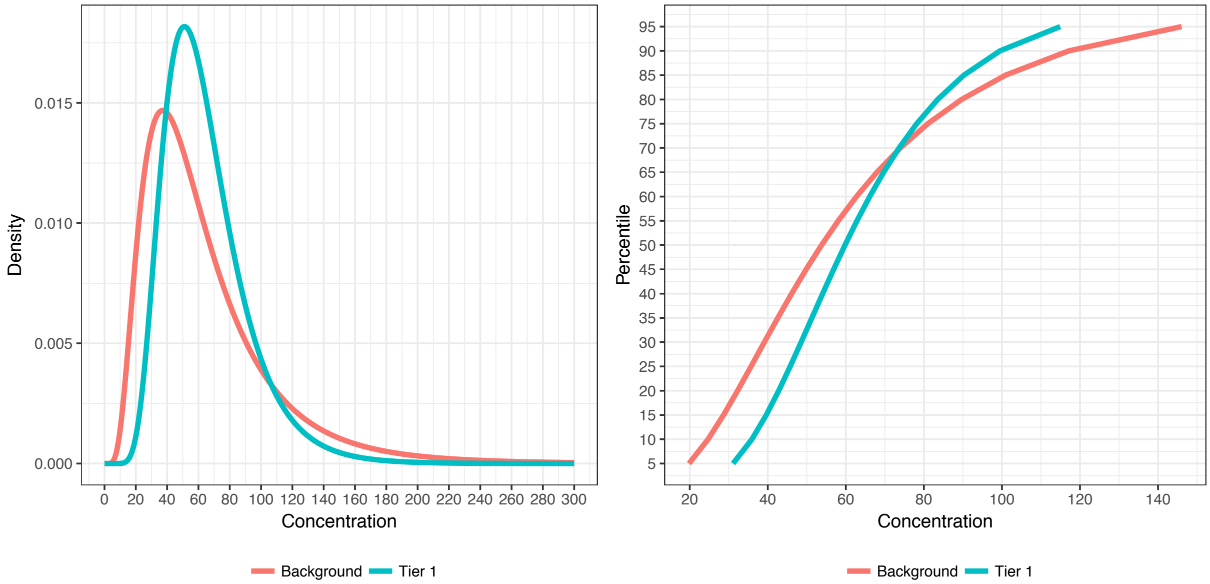

Figure A1. Fitted density and cumulative density plots for the between-property distributions estimated by the random effects models on the $\log_{10}$ scale. Normality on the $\log_{10}$ scale is assumed. Horizontal axis: mg/kg.

In order to address the larger measurement errors for properties with high lead concentrations (on the original scale) we considered the random effects model on the $\log_{10}$ scale. Statistically, the heteroscedasticity can be visually assessed using plots of the fitted values vs. the residuals (Figure A2). Figure A2 shows that for the models on the original scale (left side of Figure A2) the residuals for small fitted values are less variable than for larger fitted values. While not exactly constant, the residuals for the models on the $\log_{10}$ scale (right side of Figure A2) are relatively similar across different fitted values. This means that this type of heteroscedasticity can be addressed by the log transformation. Therefore, the between-property distributions and the confidence intervals for the difference between the Background and Tier 1 will be more valid from the $\log_{10}$ scale models than from the models on the original scale.

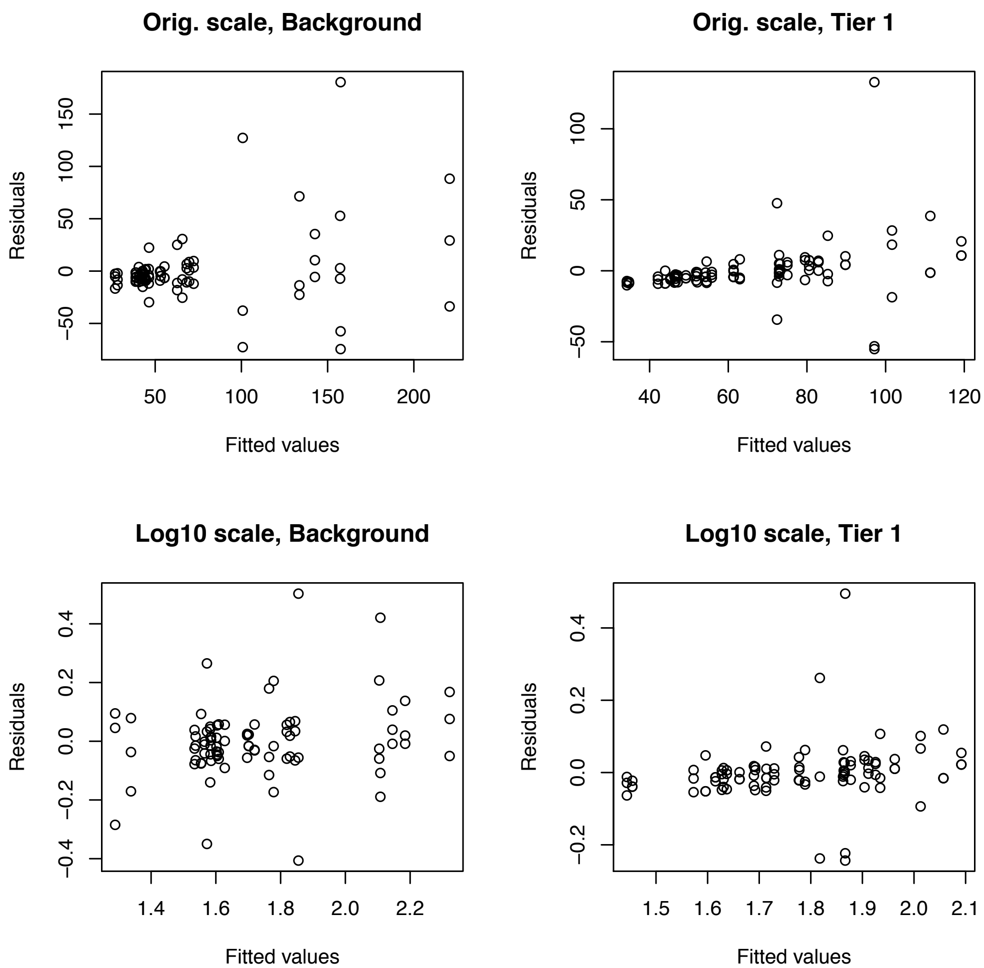

Figure A2. Fitted vs. residual heteroscedasticity diagnostics of the random effects models by scale (models on the original scale and models on the $\log_{10}$ scale) and by area (Background and Tier 1).

The formulation of the random effects models (and, consequently, the validity of the confidence intervals) also relies on the normal ("bell-shaped") distribution being a good approximation to certain distributions occurring in this analysis. The normal quantile-quantile (QQ) plots of the between-property random effects in Figure A3 show that the normality assumption is substantially violated for the model of the Background area on the original scale (left top of figure). Some deviations from normality are also observed for the model of the Background area on the $\log_{10}$ scale (left bottom) but the distribution on the $\log_{10}$ scale is better represented by a normal distribution than on the original scale. The normal distribution represents the Tier 1 data relatively well both on the original and on the $\log_{10}$ scale (very slightly better on the $\log_{10}$ scale).

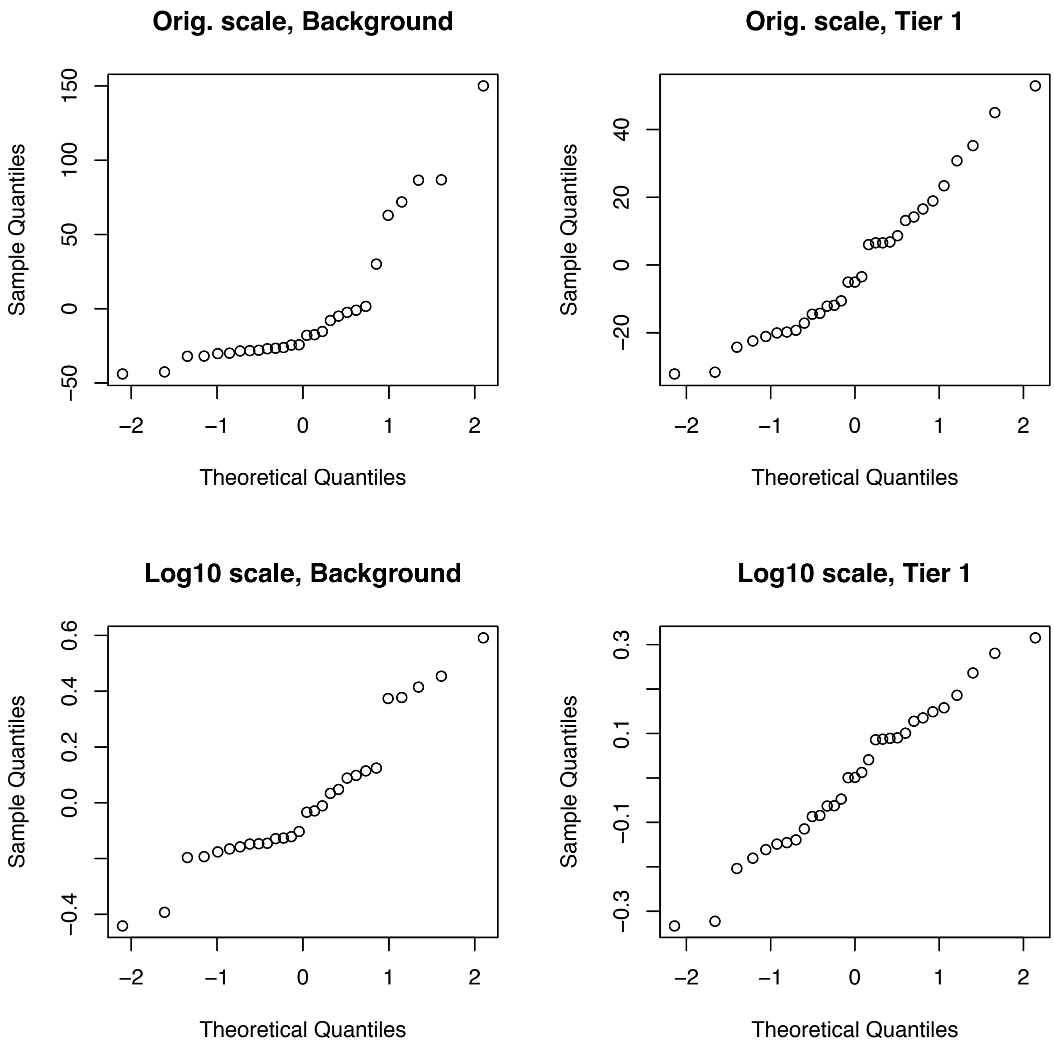

Figure A3. Normal QQ plots of between-property random effects by scale (models on the original scale and models on the log10 scale) and area (Background vs. Tier 1).

Overall, both Figures A2 and A3 suggest that the analysis on the $\log_{10}$ scale satisfies the homoscedasticity (equality of variances) and the normality assumptions better than the analysis on the original scale. For completeness, our analysis (Table 1) presents estimates and confidence intervals for models on both the original (mg/kg) and $\log_{10}$ scales. However, Figures A2 and A3 support the validity of the confidence intervals for the $\log_{10}$ models while they suggest problems with the validity of the confidence intervals for the models in the original scale. The p-values for the $\log_{10}$ model should also be more reliable than p-values from the model on the original scale.

Confidence intervals and p-values were estimated using the parametric bootstrap. Our in-house analysis also considered the profile likelihood and the non-parametric bootstrap confidence interval methods. The results for these two alternative methods are not presented because the profile likelihood method did not provide confidence intervals for the T1-B difference (as two separate random effects models were fitted to the two areas) and the non-parametric bootstrap method resulted in overoptimistically narrow confidence intervals (a consequence of the small sample size). Mathematically, the parametric bootstrap method (results shown in Tables 1 and 2) was the most reliable confidence interval method for these data.

Comments on Stantec's use of the Mann-Whitney test and on the interpretation of p-values

1. The problem with the Mann-Whitney test is that it does not compare means in these data. Instead, the Mann-Whitney tests for "stochastic equality" between two groups. Namely, the null hypothesis is that P(X>Y) equals 0.5. The alternative hypothesis is that P(X>Y) does not equal 0.5, meaning that random samples comparing a single value from each group (two values total) would find one specific group having the larger of the two values in more than half of such comparisons. In common language, one might say that members of one group are generally larger than members of the other group. For two distributions with different shapes the p-value from the Mann-Whitney test reflects not only the difference in means between the two groups but, also, differences in other features of the two distributions—embodied in their different shapes. (See Figure A1 in this Appendix that shows that there is very likely a difference in shapes of the two distributions.) An additional problem with the Mann-Whitney test for these data is that it does not adjust for the measurement error.

2. It is important to point out that null hypothesis tests of no difference (such as the Mann-Whitney test or our random effects model tests) are also of limited utility to assess <u>equivalence</u> in the context of this study. That is, even if the Mann-Whitney test was valid, a statistically non-significant p-value would not imply equality of means. Stantec appeared to interpret the statistically non-significant difference (Mann-Whitney p-values from 0.18 to 0.23) between the Tier 1 and Background lead concentrations as support for their statement that the lead concentrations "did not significantly differ." This is an incorrect interpretation of a statistically non-significant p-value. Depending on the analysis, the sample size and other factors, a statistically non-significant difference can be a difference that is large enough to be of practical importance. The p-value is often misunderstood to be the probability that two means are equal. That is not the case. Instead, the p-value is the frequency of observing results more extreme than in the actual data <u>if</u> the same study process (e.g. the same sample sizes, the same underlying distributions, etc.) was replicated <u>and</u> if the null hypothesis were true. Overall, the p-value is a frequently used statistical index that expresses strength of evidence, but it is of little use in assessing equivalence in this study.

The analyses of lead concentrations presented in this review report are based on the data in Table A2.

Table A2. Lead concentration data (mg/kg), replicates #1-3, for Background and Tier 1.

Background				Tier 1			
Property	***#1***	***#2***	***#3***	***Property***	***#1***	***#2***	***#3***
BG-1	40.3	16.7	68.8	**5**	150	110	110
BG-2	52.1	52.7	43.7	**6**	36	33	38
BG-3	187	309	250	**7**	26	27	26
BG-4	100	83	338	**8**	42	230	44
BG-5	75.7	60.2	81.9	**9**	84	69	71
BG-6	36.6	33.4	35.8	**10**	44	41	39
BG-7	42.1	42.9	27.7	**11**	38	120	64
BG-8 (*)	150	160	210	**13**	39	43	44
BG-8 (**)	189	372	1370	**14**	61	46	47
BG-9	10.1	21.6	24.2	**15**	24	26	27
BG-10	37.3	46.4	35.4	**16**	50	53	50
BG-11	57.8	96.4	40.3	**18**	130	120	83
BG-12	63	28.1	228	**19**	130	140	130
BG-13	39.9	38.8	33	**21**	42	43	38
BG-14	26.1	20	14.7	**22**	35	44	35
BG-15	87.9	51.4	44.6	**23**	57	58	71
BG-16	111	120	205	**24**	40	40	39
BG-17	137	153	178	**25**	72	74	78
BG-18	38.3	36.9	41.1	**26**	72	81	79
BG-19	48.3	34.4	42.5	**27**	44	51	50
BG-20	48.5	48.4	53	**28**	51	48	45
BG-21	30.2	30.1	44.4	**29**	83	110	78
BG-22	35.8	33.3	29.7	**30**	57	57	66
BG-23	35.8	36.2	45.5	**31**	53	55	51
BG-24	48.8	59.8	49	**33**	57	61	62
BG-25	71	57.5	75	**34**	78	74	73
BG-26	32.3	28.6	37.4	**35**	94	100	94
BG-27	78.4	59.8	70.4	**38**	89	87	73
BG-28	34.6	37	39.4	**39**	44	46	44
				40	81	84	88
*** New data for BG-8, ** Original data for BG-8**				**41**	83	90	89

Point 8:

The following plots are related to the issue of spatial correlation among lead concentrations. See the text under Point 8 in the main body of this report.

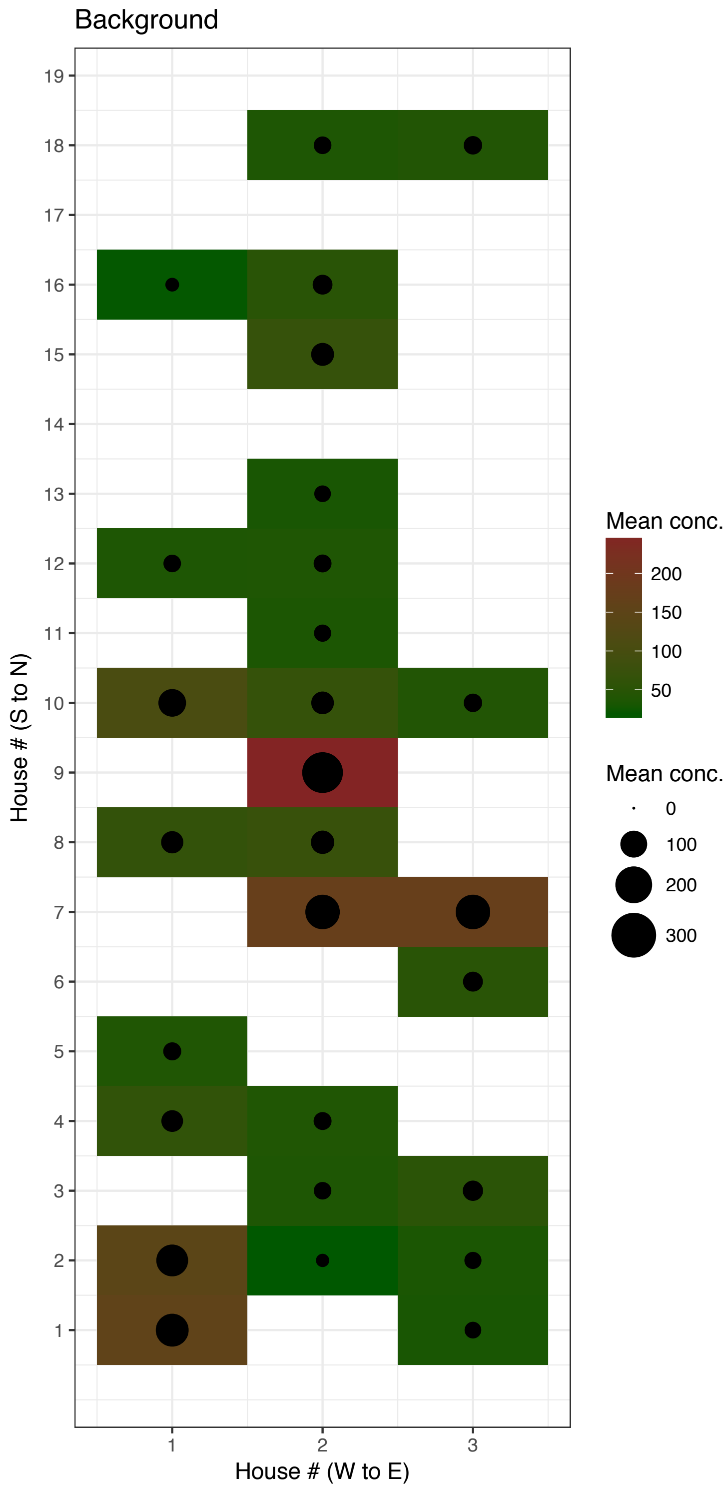

Figure A4. Spatial distribution of the three-replicate mean per property. Background area.

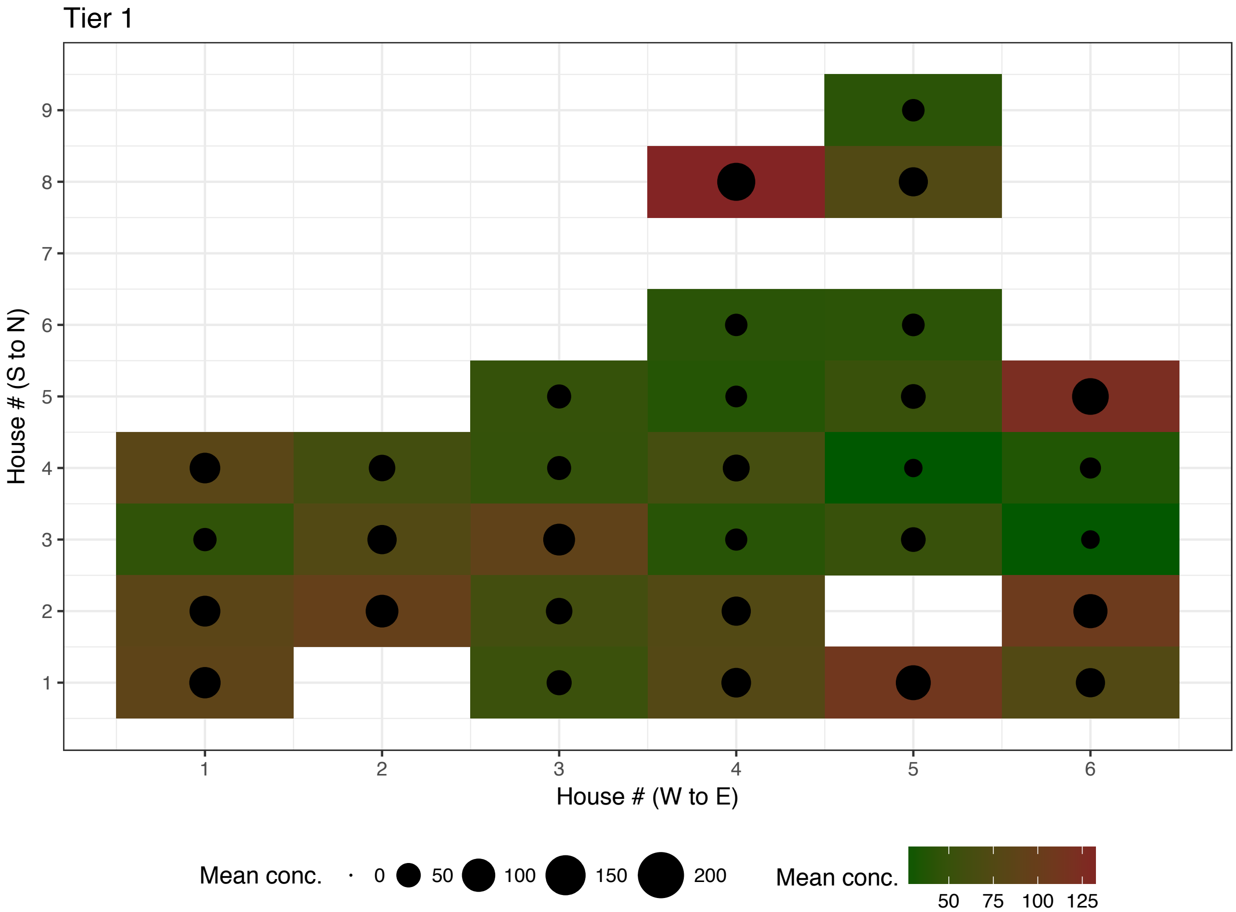

Figure A5. Spatial distribution of the three-replicate mean per property. Tier 1 area.

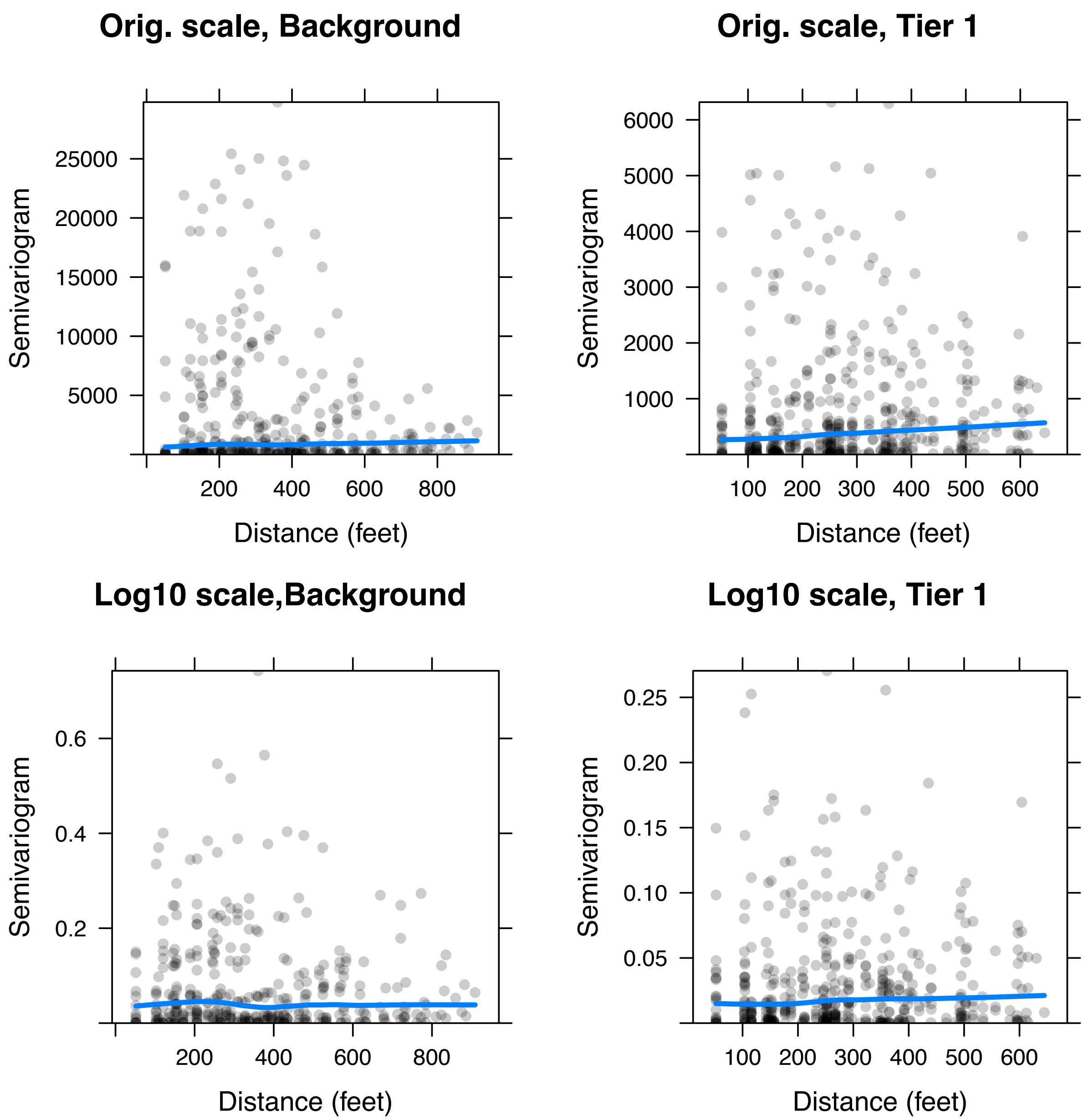

Figure A6. Variograms of the three-replicate means for the Background and Tier 1 areas. Original and $\log_{10}$ scale. Blue line = trend in mean covariance estimated by the LOESS smoother (Cleveland, Grosse, & Shyu, 1992).

The variogram shows the relationship between a) a measure of the variance in lead concentrations between a pair of properties (vertical axis, customarily labelled as "Semivariogram") and b) the distance between the two properties in a pair (horizontal axis). All possible pairs of properties are represented in each plot (Tier 1 and Background appear in separate plots). The distances used in Figures A6 and A7 are approximate. The distances were calculated from manual measurements made from Figures D1 and D2 of Stantec 2017; the conversion of manually measured distances to feet was based on the 150-foot scale presented in Figures D1 and D2 of Stantec 2017 (Stantec Consulting Services, Inc., 2017).

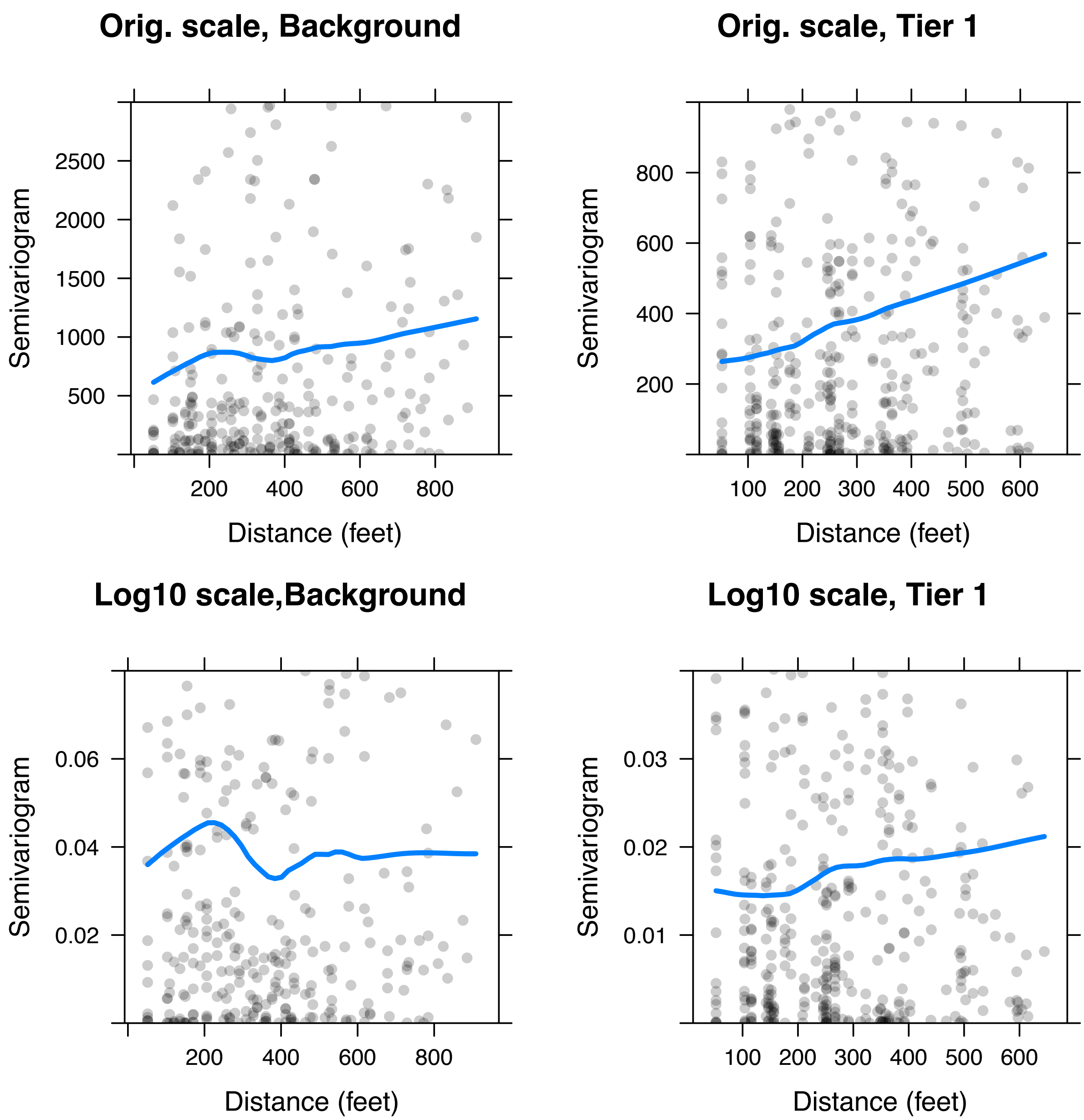

Figure A7. A modified version of Figure A6 with restricted vertical axes. Variograms of the three-replicate means for the Background and Tier 1 areas. Original and $\log_{10}$ scale. Blue line = trend in mean covariance estimated by the LOESS smoother (Cleveland, Grosse, & Shyu, 1992).

Figure A6 shows a nearly flat fitted LOESS summary line, indicating a negligible correlation of lead concentrations versus distance. Figure A6 (which expands the lower part of the plots in Figure A7) shows a small trend, visually, but, compared to the much larger full vertical scale in Figure A6, the trend is negligible.

Points 9, 10 and conclusion: no additional material.