list

ATTACHMENT C

Use of the Land H-Statistic for Calculating EPCs

The Human Health Risk Assessment (HHRA) uses three methods to calculate 95% upper confidence limits (UCLs or 95% UCLs) on the sample means for use as exposure point concentrations (EPCs) (see Vol. IIIA, pp. 4-17 to 4-21; Vol. IV, pp. 4-18 to 4-20):

- For data determined by statistical tests to be normally distributed, it uses the upper bound based on Student's t-statistic;
- For data determined by statistical tests to be lognormally distributed, it uses Land's H-statistic; and
- For data determined not to fit either distribution, it uses Hall's bootstrap sampling to predict the UCL.

GE's evaluation of these statistical techniques is provided below. As discussed, GE's principal concern is with the application of the Land H-statistic, particularly for small data sets (i.e., those with a sample size less than 30).

Student's t-Procedure

UCLs based on the Student's t-procedure are known to be robust if the data are approximately normal. Thus, if the data pass a normality test, this method should work well.

Land's H-Statistic

In its Direct Contact Risk Assessment, EPA has used the Land H-statistic to calculate the 95% UCLs for PCBs for the majority of Exposure Areas (EAs) downstream of Woods Pond and for several of the sediment EAs (HHRA, Vol. IIIA, Tables 4-3 and 4-4). In its Fish and Waterfowl Consumption Assessment, EPA has used that procedure to calculate the 95% UCLs for use as EPCs for total PCBs and PCB congeners for several fish species and reaches (largemouth bass and brown bullhead in Reaches 5 and 6, all species at Rising Pond, and smallmouth bass in Connecticut) and for duck breasts (HHRA, Vol. IV, Tables 2-18, 2-20, 2-21, 2-22, and 2-26).

EPA guidance (EPA, 2002) suggests that the H-statistic procedure developed by Land (1975) can be used to develop the 95% UCL if the data are lognormal, but warns that this method may yield upper bounds that are much too large if the data are not lognormal. For this reason, the EPA (2002) guidance recommends that the data always be tested for lognormality, as EPA did in this case using the Shapiro-Wilks Test for sample sizes < 50 and the Lilliefors Test Statistic for samples $\geq$ 50. However, the difficulty in this approach is that the distribution of environmental sampling data is often not truly lognormal and yet may appear lognormal on statistical tests. Even small deviations between the assumed lognormal distribution of the data and the actual population distribution can greatly influence the results of the H-statistic, yielding upper bounds that are much too large (Singh et al., 1997, 1999; Ginevan and Splitstone, 2002). For example, Ginevan and Splitstone (2002) have shown that data sets composed of mixtures of lognormal distributions, which usually test as not significantly different from lognormal, can still cause the Land procedure to yield very badly biased upper bounds. They showed that, in such cases, the Land H-statistic can overestimate the true upper bound by a factor of 800, and that overestimates on the order of 30- to 50-fold were common.

The problems with using the Land H-procedure for environmental data (even when they appear to be lognormally distributed) were demonstrated in another EPA guidance document (Singh et al., 1997), which is cited in EPA (2002). In that document, the authors reported the results of their evaluation and comparison of UCLs calculated through different statistical methods using a variety of sampling data sets, and noted that the UCLs obtained using the H-statistic were consistently larger and less accurate that the UCLs calculated using other statistical approaches. They stated that "it is observed that the H-UCL becomes orders of magnitudes higher even when the data were obtained from a lognormal population and can lead to incorrect conclusions. This is especially true for samples of smaller sizes (e.g., < 30)." As a result, this guidance recommends that "in environmental applications, the use of the H-UCL to obtain an estimate of the upper confidence limit of the mean should be avoided." Similar conclusions were reached by EPA consultants Schultz and Griffin (1999) in an analysis of hazardous waste sites in EPA Region 8. They noted that the H-statistic "may overestimate the exposure point concentration and may lead to unnecessary cleanup of hazardous waste sites." Moreover, EPA's 2002 guidance itself states that when sample sizes are small (less than 30), the Land H-procedure is "impractical even when the underlying distribution is lognormal."

In short, it is widely recognized that when a data set departs even slightly from a true lognormal distribution, even when it tests as lognormal, or when a data set is small (less than 30) even if lognormal, the Land H-procedure can greatly overstate the true upper bound on the mean. Thus, for environmental data sets, which are seldom truly lognormal, it follows that the Land H-procedure should not be used to calculate the 95% UCL. This is particularly true for the smaller data sets in the HHRA (i.e., those with less than 30 samples), which include: (a) the data sets in the Direct Contact Assessment for EAs for which EPA has used the Land H-statistic to calculate the 95% UCL, all but one of which have 13 or fewer data points and in most cases have less than 10; and (b) some of the data sets in the Fish and Waterfowl Consumption Assessment, such as the yellow perch data for Rising Pond (n = 14) and all the duck breast data (n = 25). At a minimum, GE believes that, consistent with EPA guidance (EPA, 2002; Singh et al., 1997), the Land H-statistic should not be used for these small data sets.

Hall's Bootstrap

Hall's bootstrap is essentially a bootstrap procedure that uses a transformation to correct for bias and skewness. Review of Hall's papers (cited in Attachment 4 to the HHRA) and Manley (1997) (also referenced in Attachment 4 to the HHRA) suggests that this procedure should work well, regardless of the underlying distribution.

Comparison of Land's H-Statistic and Hall's Bootstrap Procedure

To evaluate the performance of the Land's H-statistic and Hall's bootstrap, a simulation was performed to compare the results of Hall's bootstrap procedure with results generated using Land's H-statistic. In this simulation, the same distribution considered by Ginevan and Splitstone (2002) was used. The distribution is a mixture of four lognormal distributions, each with a geometric standard deviation of 4.5 (about 1.5 in natural logs) and geometric means of 0.01, 0.1, 1.0, and 10.0. While this is a theoretical example, a mixture of distributions would not be unexpected in the cases of environmental releases. In the floodplain, for instance, certain areas may be more subject to sediment deposition than others and therefore may be expected to have higher PCB concentrations. It was further assumed sampling occurred on a grid where 1/4 of the environment was in each distribution. Samples of size 40 with 10 observations randomly drawn from each distribution were considered. In this simulation, 5000 random samples were drawn in this manner. For each sample, the Land H 95% UCL and the Hall's bootstrap 95% UCL were calculated. The Hall's bootstrap was based on 1000 bootstrap replications per sample.

Table 1 displays a broad range of summary statistics for the these two UCL procedures as well as summary statistics for the means of the 5000 random samples, while Table 2 provides the estimate of the mean at various percentiles of the distributions.

Table 1: Summary Statistics for Simulations of UCL Procedures.

	Land UCL	**Hall UCL**	**Arithmetic Mean**
n	5000	5000	5000
St. Deviation	1270	169	6.8
Mean	662	52	8.6
Minimum	12	2.2	1.3
1st Quartile	167	12	4.8
Median	330	23	6.8
3rd Quartile	703	45	10
Maximum	33052	4376	110
Coverage	100%	85.26%	-

Table 2: Distribution Percentiles

Percentile	**Land UCL**	**Hall UCL**	**Arithmetic Mean**
5th	64	5.7	3.0
10th	92	7.3	3.5
15th	115	8.7	4.0
95th	2,226	136	20

There are a number of criteria that can be used to judge bounds. These include the degree of coverage, bias, and variability, which are discussed below. The overall weight of the various criteria can be used to select the better performing procedure for this trial case.

Coverage has been stressed in many discussions of bounds, including EPA (2002) and HHRA-Attachment 4. Coverage simply means that the estimated 95% UCL should be greater than the true mean a large majority of the time (typically 95 percent is assumed to be good). Coverage for the UCL estimators evaluated is 100 percent for Land and 85.26 percent for Hall's bootstrap. The former is much too conservative – i.e., it achieves 100 percent coverage by over-predicting the mean substantially even in the lower percentiles of the distribution of estimates, as shown in Table 2. While Hall's coverage is less than 95 percent, the 5th percentile of the Hall estimator is 5.7, which is about 2/3 of the true mean value, and the 10th percentile is 7.3, which is about 85% of the true mean value. That is, while 15 percent of the Hall estimates are technically too low, they do not miss the true mean by a large amount, as was the case for the Land procedure. Thus, UCLs calculated using Hall's bootstrap are preferable in terms of coverage.

Another way of looking at any statistical estimator is bias. That is, we would like an estimator to have an expected value equal to the true expectation of the statistic of interest. For our distribution and samples of size 40, the 95th percentile is about 20. The mean UCL using the Land H and Hall's bootstrap procedures are 662 and 52, respectively, while the median UCLs are 330 and 23. Here, the Land estimates are extremely high for both mean and median UCLs, while the Hall's bootstrap statistic is a somewhat high from a mean perspective and almost perfect from a median perspective. Here again the Hall's bootstrap appears to be the better performer.

Finally, an estimator should have small variability. The standard deviation (as shown in Table 1) is 1270 for Land H and 169 for Hall's bootstrap. Here again the Hall's procedure appears to be better.

Thus, based on the weight of all criteria, the Hall's bootstrap procedure for estimating the 95% UCL is superior for the test distribution used in this example. Yet this distribution would be unlikely to be ruled out as a lognormal distribution using standard methods of evaluating distribution fits.

EPA has shown (in the HHRA-Attachment 4) that, from a coverage perspective, Hall's bootstrap does well for a variety of distributions. Those simulations indicate that the procedure also performs well with respect to minimization of bias and reasonable variability. While Hall's bootstrap did not do as well in our simulation as it did in those performed by EPA, the distribution involved in our simulation was deliberately contrived to pose difficult problems. The difference in the test provided here *vis-a-vis* the HHRA-Attachment 4 test distributions is that Hall's bootstrap proved to be superior to the H-statistic for a sample distribution that, in all likelihood, would have been found to be consistent with a lognormal distribution and thus evaluated using the Land procedure under EPA's approach. While GE does not disagree that distribution fit testing is appropriate in most cases, we conclude that the H-statistic is so biased and so sensitive to the assumption of lognormality that standard procedures of testing for distribution fit may not avoid unreasonable estimates of the UCL. Again, as discussed above, this is particularly true for samples of smaller sizes. Since Hall's bootstrap is shown to perform reasonably well even for the extreme test distribution, it would be appropriate to expand its application to any non-normally distributed data sets.

Summary

Of the three procedures used in the HHRA to calculate the 95% UCL, the Land H-statistic has poor properties and can cause substantial overestimates of the true upper bound when used with environmental data, even if the data distributions test as lognormal. As recognized in EPA guidance (EPA, 2002; Singh et al., 1997), this is particularly true for sample sizes less than 30. GE believes that the Land H-procedure should not be used at all in the HHRA to estimate EPCs. At a minimum, however, that procedure should clearly not be used for data sets with less than 30 samples. Instead, GE recommends that EPA use Hall's bootstrap procedure to calculate the 95% UCLs for such data sets that are not normally distributed.

References

EPA. 2002. *Calculating Upper Confidence Limits for Exposure Point Concentrations at Hazardous Waste Sites.* U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response. OSWER 9285.6-10.

Ginevan, M.E. and D.E. Splitstone. 2002. Bootstrap upper bounds for the arithmetic mean of right-skewed data, and the use of censored data. *Environmetrics* 13: 443-464.

Land, C.E. 1975. Tables of confidence limits for linear functions of the normal mean and variance. In: *Selected Tables in Mathematical Sciences*, Vol. III. American Mathematical Society, Providence RI. Pp. 385-419.

Schultz, T.W., and S. Griffin. 1999. Estimating risk assessment exposure point concentrations when the data are not normal or lognormal. *Risk Analysis* 19:577-584.

Singh, A.K., A. Singh, and M. Engelhardt. 1997. *The Lognormal Distribution in Environmental Applications.* U.S. Environmental Protection Agency. EPA/600/R-97-006. December.

Singh, A.K., A. Singh, and M. Engelhardt. 1999. *Some Practical Aspects of Sample Size and Power Computations for Estimating the Mean of Positively Skewed Distributions in Environmental Applications.* U.S. Environmental Protection Agency. EPA/600/s-99/006. November.