Appendix D

Comparison of ProUCL and Default Data Distributions for Calculation of Exposure Point Concentrations and Risks

APPENDIX D. COMPARISON OF PROUCL AND DEFAULT DATA DISTRIBUTIONS FOR CALCULATION OF EXPOSURE POINT CONCENTRATIONS AND RISKS

D.1 Introduction

Historically, US Environmental Protection Agency (USEPA) risk assessment protocols have assigned data sets of sufficient quantity (taken as at least ten data points for this HHRA) to either a normal or lognormal distribution (USEPA, 1989). (For smaller data sets, the maximum detected value is used as the exposure point concentration [EPC].) This Onondaga Lake HHRA currently uses a "best-fit" approach to assign a normal or a lognormal distribution to each data set and subsequently to calculate the 95 percent upper confidence limit (UCL) on the mean. (The 95 percent UCL is used as an estimator of the arithmetic mean of the data set, and is used as the EPC for risk calculations, using equations specific to either normal or lognormal distributions [USEPA, 1992d].) However, USEPA has recently been moving toward using non-parametric tests to calculate the UCL on the mean where the data set does not fit either a normal or lognormal distribution. USEPA Las Vegas has developed software (ProUCL 2.1; USEPA, 2002c) that performs the statistical tests to determine the data distribution type (i.e., tests for normal and lognormal distribution) and also calculates the UCL by a number of different statistical techniques.

ProUCL software calculates multiple 95 percent UCLs on the mean based on several statistical methods. The distribution and skew of any given data set dictate which of the values calculated by the software should be applied. Unless otherwise noted, the 95 percent UCLs for the lognormal data sets were calculated using Land's H statistic (USEPA, 1992). The non-parametric UCLs were estimated using the Chebyshev (mean, std) method. Depending on the skew of the data set, the EPC may be estimated with the 95 percent Chebyshev (mean, std) UCL, the 97.5 percent Chebyshev (mean, std) UCL, or the 99 percent Chebyshev (mean, std) UCL (for discussion see USEPA, 2002]). Unless otherwise noted, the 95 percent Chebyshev (mean, std) UCL was used to estimate the non-parametric EPCs.

To assess the potential impact of the initial forcing of all data into either normal or lognormal distribution, a number of the data sets (focusing on contaminants of potential concern [COPCs] responsible for a significant percent of the overall risks calculated for individual matrices) were run through USEPA's ProUCL software. The purpose of this was to:

- Confirm the distribution of the data.
- Recalculate, if necessary, the 95 percent UCL on the mean (and corresponding EPC) based on the new (i.e., non-parametric) distribution.
- Recalculate the associated cancer risk or non-cancer hazard based on new UCL/EPC calculations.

D.2 Data Distribution

Previous calculations had determined, using the best-fit method, that the majority of the data for all matrices for which there are at least ten samples (which is considered the minimum number of samples for usable statistical analysis) had a lognormal distribution. Running the data through ProUCL confirmed that the majority of the data sets reviewed, which were initially identified as lognormal (11 out of 16), do fit a lognormal distribution (Table D-1). One data set initially evaluated as normal and one data set for which the maximum value had been used as the EPC were also assessed by ProUCL and found to have non-parametric distributions for which the ProUCL software calculated UCLs.

Except as noted specifically in the discussion in Section D.3 below, the data in all three matrices assessed (fish, northern basin sediments, and southern basin sediments) were determined by ProUCL to be lognormally distributed, confirming the assumption previously made; therefore, those EPCs are the same as those calculated previously.

D.3 Recalculation of UCLs using ProUCL 2.1

D.3.1 Scope and Rationale of Data Sets Selected for Recalculation

The fish tissue and the lake bottom (northern basin and southern basin) sediment pathways were evaluated. It was not necessary to perform this analysis on the wetland or dredge spoils data because there are fewer than ten data points available. As risks and hazards associated with the surface water pathway were low (i.e., hazard index [HI] less than 1.0, and cancer risk less than 10^{-6}), the surface water data were not recalculated. For all matrices where the chemical was not detected, half of the quantitation limit was used as the concentration. This convention has been used consistently throughout this project. The COPCs evaluated are listed in Table D-1 by matrix.

D.3.2 Fish Tissue

Using ProUCL 2.1, mercury and PCB fish tissue data sets were determined to have non-parametric distributions. (ProUCL 2.1 confirmed the initial assignment of lognormal distributions to the arsenic and PCDD/PCDF [toxicity equivalence quotient, or TEQ] data sets.) The non-parametric 95 percent UCLs on the mean calculated were similar to (within 10 percent of) the original UCLs (Table D-1). The non-parametric UCL on the mean for mercury is 1.13 mg/kg, slightly higher than the original UCL, based on a normal distribution, of 1.08 mg/kg. PCBs have a non-parametric (97.5 percent Chebyshev [mean, std]) UCL of 1.0 mg/kg versus the original lognormal 95 percent UCL on the mean of 0.91 mg/kg.

D.3.3 Sediment – General

For sediment, the 0 to 0.3 meter (m) depth samples or the length-weighted average of samples in this interval were used to calculate the 95 percent UCLs on the mean.

D.3.4 Northern Basin Sediments

In the northern basin sediments, antimony, benzo(a)pyrene, and hexachlorobenzene were determined by ProUCL to have non-parametric distributions, and the resulting EPCs were noticeably different (Table D-1). The EPCs for antimony (97.5 percent Chebyshev [mean, std] UCL) and hexachlorobenzene (99 percent Chebyshev [mean, std] UCL) decreased, going from 9.2 to 6.78 mg/kg and 0.41 to 0.22 mg/kg, respectively. (The original lognormal UCL calculated for antimony was greater than the maximum value of 9.2 mg/kg, so the maximum was used as the EPC.)

The EPC for benzo(a)pyrene (97.5 percent Chebyshev [mean, std] UCL) went from 0.156 to 0.345 mg/kg, more than double the original EPC. It should be noted that the northern basin data set has fewer data points than the southern basin data set, with an average of 29 values (ranging from 25 to 36) for each COPC evaluated.

D.3.5 Southern Basin Sediments

Seven COPCs, consisting of two metals (arsenic and chromium), one SVOC (hexachlorobenzene), and four PAHs (benzo[a]pyrene, benzo[b]fluoranthene, dibenz[a,h]anthracene, and naphthalene) were reevaluated using the ProUCL 2.1 software. The initial assignment of a lognormal distribution was confirmed for all seven COPCs. However, the ProUCL software provides a recommendation of the UCL value to use for lognormal data sets. This recommendation is based on the degree of skew in the data, and for data sets with extremely high skew values (defined as standard deviations greater than 2.0), ProUCL no longer recommends the H-statistic-based UCL calculation (which is the USEPA default calculation method [USEPA, 1992d]).

Two of the southern basin sediments, naphthalene and hexachlorobenzene, had high skew values (standard deviations between 2.5 and 3.0), and for these the ProUCL software recommended use of the greater of the 95 percent Chebyshev (mean, std) UCL and the 95 percent Chebyshev (minimum variance unbiased estimator [MVUE]) UCL. Using the recommended UCL calculation, the ProUCL-based EPC for naphthalene approximately doubled, from 1,050 to 2,190 mg/kg; whereas the EPC for hexachlorobenzene decreased by more than half, from 10.5 to 3.8 mg/kg.

The ProUCL documentation identifies data sets with a standard deviation approaching or exceeding 3 as "badly behaved" and requiring further investigation before the EPC can be accurately estimated. The standard deviations for hexachlorobenzene (2.73) and naphthalene (2.95) approached this value. In addition, the maximum naphthalene concentration was 26,000 mg/kg, while the next highest concentration was 145 mg/kg. (When the 26,000 is removed from the data set, the EPC changes from 2,190 to 52.5 mg/kg; the skew is still high, with a standard deviation of 2.7.) Hexachlorobenzene data are lognormally distributed, but because the data has high skew, the 95 percent Chebyshev (MVUE) UCL was applied.

D.4 Recalculation of Risk and Hazard Using ProUCL-Generated Exposure Point Concentrations

The non-parametric UCLs generated using ProUCL were used to recalculate the overall risks in the fish tissue and lake bottom sediments (see Table D-2).

There are no significant changes to the overall cancer risks and non-cancer hazards in fish tissue based on the use of the non-parametric UCLs. Changes to the non-cancer HIs for fish ingestion increased only minimally (by less than 1 percent) for both the RME and CT scenarios. Minimal increases (about 1 to 2 percent) were noted for RME and CT cancer risks calculated using the ProUCL-generated EPCs.

In the southern basin sediments, the cancer risks decreased slightly as a result of the change in the hexachlorobenzene EPC. (Naphthalene has no cancer slope factor, and therefore changes to the naphthalene EPC do not impact the cancer risk.) The non-cancer hazards in the southern basin did increase based on the change in the naphthalene EPC. The most significant difference from the previous calculations is that the non-cancer hazard increased from 0.54 to 0.70 (but is still below 1.0) for young children in the southern basin.

The overall cancer risks calculated in the northern basin were slightly higher (because of the increase in benzo[a]pyrene EPC) and non-cancer hazards were slightly lower (due to a decrease in the antimony and hexachlorobenzene EPCs), but the change in these values did not affect whether or not there was a risk exceeding target levels for a particular pathway.

D.5 References

US Environmental Protection Agency (USEPA). 1989. Risk Assessment Guidance for Superfund, Volume I, Human Health Evaluation Manual (Part A); Section 6.4. EPA/540/1-89/002. December. Interim Final.

USEPA. 1992. Supplemental Guidance to RAGS: Calculating the Concentration Term. USEPA Office of Solid Waste and Emergency Response Publication 9285.7-08I (NTIS number PB-92-963373). May.

USEPA. 2002. ProUCL Version 2.1. Software and User's Guide. Prepared for USEPA Region 3 by Lockheed-Martin Environmental Systems and distributed by USEPA National Exposure Research Laboratory, Las Vegas, NV. July.

Table D-1. Exposure Point Concentrations for Key COPCs Evaluated by ProUCL 2.1 (listed by matrix)

FISH

	HHRA Default Values (distribution forced to normal or lognormal)		Values and distribution calculated by ProUCL 2.1 software				General Statistics	
Chemical of Potential Concern	Best-Fit UCL (mg/kg)	Best-Fit Distribution (normal, lognormal, or maximum)	ProUCL (mg/kg)	ProUCL Distribution	Standard deviation (of logtransformed variable)	Skew	Percent Difference [(ProUCL/Original) x 100%]	Number of data points (n)
Arsenic	8.0E-01	lognormal	8.0E-01	lognormal	0.95	mild to moderate	**-0.1%**	11
Mercury (as MeHg)	**1.08E+00**	**normal**	**1.13E+00**	**non-parametric**	**0.56**	**mild to moderate**	**4.6%**	**728**
PCBs[1]	**9.13E-01**	**lognormal**	**1.00E+00**	**non-parametric**	**1.01**	**moderate to high**	**9.7%**	**130**
High MW PCBs[1]	**4.79E-01**	**lognormal**	**4.52E-01**	**non-parametric**	**1.35**	**moderate to high**	**-5.6%**	**120**
Low MW PCBs[1]	**5.77E-01**	**lognormal**	**6.25E-01**	**non-parametric**	**1.03**	**moderate to high**	**8.4%**	**130**
PCDD/PDCF (TEQ-Total)	1.96E-05	lognormal	1.96E-05	lognormal	1.26	moderate to high	**-0.2%**	30

NORTHERN BASIN SEDIMENT

	HHRA Default Values (distribution forced to normal or lognormal)		Values and distribution calculated by ProUCL 2.1 software				General Statistics	
Chemical of Potential Concern	Best-Fit UCL (mg/kg)	Best-Fit Distribution (normal, lognormal, or maximum)	ProUCL (mg/kg)	ProUCL Distribution	Standard deviation (of logtransformed variable)	Skew	Percent Difference [(ProUCL/Original) x 100%]	Number of data points (n)
Arsenic	3.86E+00	lognormal	3.86E+00	lognormal	1.11	moderate to high	**0.0%**	25
Antimony[1]	**9.20E+00**	**Max concentration**	**6.78E+00**	**non-parametric**	**1.93**	**moderate to high**	**-26.3%**	**25**
Iron	5.79E+03	lognormal	5.79E+03	lognormal	0.72	mild to moderate	**0.1%**	25
Benzo (a) pyrene[1]	**1.56E-01**	**lognormal**	**3.45E-01**	**non-parametric**	**1.12**	**moderate to high**	**121.2%**	**36**
Hexachlorobenzene[3]	**4.06E-01**	**lognormal**	**2.21E-01**	**non-parametric**	**2.14**	**extremely high**	**-45.6%**	**32**

SOUTHERN BASIN SEDIMENT

	HHRA Default Values (distribution forced to normal or lognormal)		Values and distribution calculated by ProUCL 2.1 software				General Statistics	
Chemical of Potential Concern	Best-Fit UCL (mg/kg)	Best-Fit Distribution (normal, lognormal, or maximum)	ProUCL (mg/kg)	ProUCL Distribution	Standard deviation (of logtransformed variable)	Skew	Percent Difference [(ProUCL/Original) x 100%]	Number of data points (n)
Arsenic	1.24E+01	lognormal	1.24E+01	lognormal	1.15	moderate to high	**0.0%**	51
Chromium	2.05E+02	lognormal	2.05E+02	lognormal	1.32	moderate to high	**0.0%**	66
Hexachlorobenzene[2]	**1.05E+01**	**lognormal**	**3.83E+00**	**lognormal**	**2.73**	**extremely high**	**-63.5%**	**63**
Benzo (a) pyrene	7.36E+00	lognormal	7.36E+00	lognormal	1.97	high	**0.0%**	62
Dibenz (a,h) anthracene	1.62E+00	lognormal	1.62E+00	lognormal	1.89	high	**0.0%**	62
Benzo (b) fluoranthene	9.75E+00	lognormal	9.75E+00	lognormal	1.93	high	**0.0%**	62
Naphthalene[2]	**1.05E+03**	**lognormal**	**2.19E+03**	**lognormal**	**2.95**	**extremely high**	**108.8%**	**64**

Notes:

HHRA default values as shown on RAGS Tables 3.1, 3.2, and 3.3 except as noted. Data distribution assigned using best-fit methodology as described in HHRA Section 4.7.3.

Highlighting (and bold text) indicates UCL changed based on the determination of the data distribution.

Unless otherwise noted, lognormal UCLs were calculated with Land's H Statistic and non-parametric UCLs used 95% Chebyshev(mean,std) UCL; for detailed discussion of UCL selection see the ProUCL manual, "Recommendations to Compute a 95th Percentile UCL of the Population Mean."

* EPC values from RAGS Table 7.1.

[1] Chebyshev (mean, std) 97.5% UCL used to estimate the Exposure Point Concentration (EPC).

[2] ProUCL software recommendation is based on skew and sample size; the Land's H Statistic was determined not to be the appropriate estimator of the 95%UCL, the greater of the 95% Chebyshev (MVUE) UCL and 95% Chebyshev (mean, std) UCL was used.

[3] Chebyshev (mean, std) 99% UCL used to estimate the Exposure Point Concentration (EPC).

Table D-2. Comparison of Cancer Risk and Non-Cancer Hazard Calculations using Default Distributions and ProUCL Software

	Non-Cancer Hazard Index				Cancer Risk			
	RME		Central Tendency		RME		Central Tendency	
Pathway	Original	ProUCL Updated	Original	ProUCL Updated	Original	ProUCL Updated	Original	ProUCL updated
Fish Ingestion - Adult Angler	**18.2**	**19.1**	**4.48**	**4.70**	**7.8E-04**	**8.1E-04**	**4.3E-05**	**4.4E-05**
Fish Ingestion - Young Child Angler	**28.3**	**29.8**	**6.97**	**7.31**	**2.4E-04**	**2.5E-04**	**4.4E-05**	**4.5E-05**
Fish Ingestion - Older Child Angler	**19.8**	**20.8**	**4.86**	**5.10**	**3.4E-04**	**3.5E-04**	**4.6E-05**	**4.7E-05**
Sediments - Northern Basin - Adult Recreational	0.020	0.019	0.007	0.007	**1.3E-06**	**1.5E-06**	1.4E-07	1.7E-07
Sediments - Northern Basin - Young Child Recreational	0.221	0.217	0.060	0.058	**3.8E-06**	**4.7E-06**	5.7E-07	6.7E-07
Sediments - Northern Basin - Older Child Recreational	0.070	0.071	0.012	0.012	**3.9E-06**	**5.0E-06**	2.5E-07	3.1E-07
Sediments - Northern Basin - Construction Worker	0.037	0.036	0.013	0.013	1.5E-07	1.8E-07	3.8E-08	4.4E-08
Sediments - Southern Basin - Adult Recreational	0.039	0.048	0.007	0.009	**1.0E-05**	**9.6E-06**	5.3E-07	5.1E-07
Sediments - Southern Basin - Young Child Recreational	0.535	0.704	0.047	0.056	**3.2E-05**	**3.1E-05**	**2.0E-06**	**1.9E-06**
Sediments - Southern Basin - Older Child Recreational	0.253	0.352	0.012	0.016	**3.5E-05**	**3.4E-05**	**1.0E-06**	9.6E-07
Sediments - Southern Basin - Construction Worker	0.219	0.273	0.062	0.072	**3.7E-06**	**3.6E-06**	8.3E-07	8.0E-07
Sediments - Wetland #6 (North) - Adult Recreational	0.042	n<10	0.015	n<10	**6.5E-05**	**n<10**	**7.1E-06**	**n<10**
Sediments - Wetland #6 (North) - Older Child Recreational	0.115	n<10	0.026	n<10	**2.6E-04**	**n<10**	**1.4E-05**	**n<10**
Sediments - Wetland #6 (North) - Construction Worker	0.078	n<10	0.029	n<10	**7.6E-06**	**n<10**	**1.5E-06**	**n<10**
Sediments - Wetland #10 (North) - Adult Recreational	0.041	n<10	0.015	n<10	**5.0E-06**	**n<10**	5.4E-07	n<10
Sediments - Wetland #10 (North) - Older Child Recreational	0.161	n<10	0.026	n<10	**1.7E-05**	**n<10**	**1.0E-06**	**n<10**
Sediments - Wetland #10 (North) - Construction Worker	0.076	n<10	0.026	n<10	6.0E-07	n<10	1.4E-07	n<10
Sediments - Wetland #12 (South) - Adult Recreational	0.023	n<10	0.004	n<10	**3.7E-06**	**n<10**	**1.9E-07**	n<10
Sediments - Wetland #12 (South) - Older Child Recreational	0.122	n<10	0.007	n<10	**1.4E-05**	**n<10**	**3.7E-07**	n<10
Sediments - Wetland #12 (South) - Construction Worker	0.135	n<10	0.042	n<10	**1.4E-06**	**n<10**	**2.7E-07**	n<10
Sediments - Wetland #19 (South) - Adult Recreational	0.027	n<10	0.005	n<10	**1.4E-05**	**n<10**	7.7E-07	n<10
Sediments - Wetland #19 (South) - Older Child Recreational	0.157	n<10	0.009	n<10	**4.9E-05**	**n<10**	**1.4E-06**	**n<10**
Sediments - Wetland #19 (South) - Construction Worker	0.156	n<10	0.047	n<10	**5.4E-06**	**n<10**	**1.2E-06**	**n<10**
Soils - Dredge Spoils (Surface) - Adult Recreational	0.026	n<10	0.009	n<10	**1.8E-06**	**n<10**	1.9E-07	n<10
Soils - Dredge Spoils (Surface) - Older Child Recreational	0.075	n<10	0.016	n<10	**4.7E-06**	**n<10**	3.5E-07	n<10
Soils - Dredge Spoils (Surface) - Construction Worker	0.048	n<10	0.018	n<10	2.1E-07	n<10	6.0E-08	n<10
Soils - Dredge Spoils (Subsurface) - Construction Worker	0.126	n<10	0.043	n<10	**1.1E-06**	**n<10**	2.4E-07	n<10
Surface Water - Adult Recreational	0.020	n<10	0.007	n<10	6.1E-08	n<10	7.8E-09	n<10
Surface Water - Young Child Recreational	0.037	n<10	0.014	n<10	2.5E-08	n<10	9.9E-09	n<10
Surface Water - Older Child Recreational	0.024	n<10	0.009	n<10	3.0E-08	n<10	9.4E-09	n<10
Surface Water - Construction Worker	0.002	n<10	0.001	n<10	4.2E-10	n<10	1.1E-10	n<10

Notes:

Hazard Indices and Cancer Risks in **boldface** exceed target levels (HI > 1, cancer risk > 10^{-6})

For data sets where N <10, the maximum concentration was used as the EPC in the HHRA. The data set is too small for meaningful statistical analysis (and no ProUCL calculation was attempted).