PORTLAND HARBOR RI

APPENDIX G
BASELINE ECOLOGICAL RISK ASSESSMENT
ATTACHMENT 7
SQG RELIABILITY ANALYSIS

DRAFT

DO NOT QUOTE OR CITE
This document is currently under review by US EPA and its federal, state, and tribal partners, and is subject to change in whole or in part.

August 19, 2009

Prepared for
The Lower Willamette Group

Prepared by
Windward Environmental LLC

TABLE OF CONTENTS

1.0 INTRODUCTION 1

2.0 RELIABILITY ANALYSIS OF GENERIC SQGS FOR INDIVIDUAL CHEMICALS 2

3.0 RELIABILITY ANALYSIS OF MEAN QUOTIENTS 4

4.0 RELIABILITY ANALYSIS OF ESBS 8

5.0 RELIABILITY ANALYSIS OF LOW SQGS 11

6.0 REFERENCES 14

LIST OF TABLES

Table 2-1. False Positive and False Negative Rates Associated with Each Set of High SQGs Based on a Comparison to High Reference Thresholds 3

Table 2-2. False Positive and Negative Rates Associated with Each Set of High SQGs Based on a Comparison to Low Reference Thresholds..... 3

Table 3-1. False Positive and Negative Rates Associated with Each High SQG Quotient Based on a Comparison to High Reference Thresholds..... 5

Table 3-2. False Positive and Negative Rates Associated with Each High SQG Quotient Based on a Comparison to Low Reference Thresholds..... 6

Table 4-1. False Negative and False Positive Rates Associated with ESB PAH Based on High Reference Thresholds 8

Table 5-1. Generic SQGs and Their Derivation and Narrative Intent..... 11

Table 5-2. Number of Stations Predicted to be Non-Toxic Based on the Sets of Low SQGs12

LIST OF MAPS

Map 2-1. Locations with no exceedances of TEC SQGs
Map 2-2. Locations with no exceedances of ERL SQGs
Map 2-3. Locations with no exceedances of TEL SQGs

1.0 INTRODUCTION

The ability of the generic sets of SQGs to accurately predict adverse effects on benthic invertebrates in Portland Harbor was evaluated by comparing results predicted from the sets of high generic SQGs with actual toxicity test results from Study Area sediments. The reliability analysis included an evaluation of both the false negative and false positive rates. This analysis was carried out for the five sets of high generic SQGs (i.e., PEC, PEL, SL2, ERM, and CSL), the associated mean quotients, and the ESBs for PAH mixtures and non-ionic organic compounds. A reliability analysis was not performed for ESBs for dieldrin and endrin because there were too few data points (endrin exceeded one SQG in one sample). In addition, the reliability of the low SQGs was evaluated using the false negative rates. This attachment presents a reliability analysis of the SQGs, mean quotients, and ESBs (Sections 2.0 through 5.0).

2.0 RELIABILITY ANALYSIS OF GENERIC SQGS FOR INDIVIDUAL CHEMICALS

The predictions of toxicity based on each set of high generic SQGs were compared to the site-specific toxicity test results for both the low and high reference thresholds derived using the reference envelope approach (Section 6.1.1.1 and Attachment 6). Similar to the reliability analysis for the site-specific SQGs, the acceptability criteria for the false negative and false positive rates were ≤ 20%. Tables 2-1 and 2-2 present the false negative and false positive rates for the five sets of high generic SQGs compared to the site-specific toxicity test results. None of the five sets of high generic SQGs could reliably predict toxicity measured as an empirical test result exceeding either the high or low reference thresholds.

Table 2-1. False Positive and False Negative Rates Associated with Each Set of High SQGs Based on a Comparison to High Reference Thresholds

SQG Set	Percentage of Toxicity Test Sampling Locations Above Generic SQG Set	*Chironomus* Mortality (%) False Positive	*Chironomus* Mortality (%) False Negative	*Chironomus* Biomass (%) False Positive	*Chironomus* Biomass (%) False Negative	*Hyalella* Mortality (%) False Positive	*Hyalella* Mortality (%) False Negative	*Hyalella* Biomass (%) False Positive	*Hyalella* Biomass (%) False Negative	All Four Endpoints Combined (%) False Positive	All Four Endpoints Combined (%) False Negative
PEC	35	32	29	30	26	32	12	33	13	30	25
PEL	54	51	17	49	11	51	0	51	0	49	11
RSET_SL2	38	35	33	34	31	35	18	36	20	33	31
ERM	41	38	29	36	23	38	12	38	13	36	22
CSL	28	26	46	24	40	26	29	26	27	24	39

CSL – cleanup screening level
ERM – effects range – median
PEC – probable effects concentration
PEL – probable effects level
RSET – Regional Sediment Evaluation Team
SL2 – screening level 2
SQG – sediment quality guideline

Table 2-2. False Positive and Negative Rates Associated with Each Set of High SQGs Based on a Comparison to Low Reference Thresholds

SQG Set	Percentage of Toxicity Test Sampling Locations Above Generic SQG Set	*Chironomus* Mortality (%) False Positive	*Chironomus* Mortality (%) False Negative	*Chironomus* Biomass (%) False Positive	*Chironomus* Biomass (%) False Negative	*Hyalella* Mortality (%) False Positive	*Hyalella* Mortality (%) False Negative	*Hyalella* Biomass (%) False Positive	*Hyalella* Biomass (%) False Negative	All Four Endpoints Combined (%) False Positive	All Four Endpoints Combined (%) False Negative
PEC	35	31	35	29	31	32	20	33	37	30	43
PEL	54	50	19	48	17	51	5	51	19	48	25
RSET_SL2	38	34	38	32	33	35	25	36	41	33	45
ERM	41	37	32	35	27	38	20	39	41	36	42
CSL	28	25	46	22	40	26	35	26	52	23	52

CSL – cleanup screening level
ERM – effects range – median
PEC – probable effects concentration
PEL – probable effects level
RSET – Regional Sediment Evaluation Team
SL2 – screening level 2
SQG – sediment quality guideline

3.0 RELIABILITY ANALYSIS OF MEAN QUOTIENTS

The predictive power of the mean quotients was assessed by evaluating the false positive and false negative rates; specifically empirical toxicity test results were compared to the mean quotient predictions of toxicity, using the thresholds of 1.0 and 0.7 as stated in EPA's Problem Formulation (Attachment 2). Tables 3-1 and 3-2 present the comparisons of the mean quotients and the adverse effects in the 293 toxicity tests; Table 3-1 uses the high reference thresholds, and Table 3-2 uses the low reference thresholds. The percentage of sampling locations predicted to have adverse effects on benthic invertebrates based on the mean quotients ranged from 6 to 15%. Compared to the toxicity test results based on the high reference thresholds, all mean quotients for the separate endpoints had low false positive rates (3.1 to 12%) and high false negative rates (24 to 69%). Combining the toxicity test results based on all four endpoints (adverse effects in any endpoint is counted as adverse effect for that sample) gave similar low false positive rates (3.1 to 9.3%) and high false negative rates (47 to 69%).

The false positive and false negative rates were similar for the low reference threshold comparison (false positive rates of 3.1 to 11% and false negative rates of 30 to 74%) for all mean quotients of the separate endpoints. By combining the toxicity test results for all four endpoints, the false positive rates remained low (between 3.4 and 8.6%) and the false negative rates increased to 62 to 82%.

Table 3-1. False Positive and Negative Rates Associated with Each High SQG Quotient Based on a Comparison to High Reference Thresholds

SQGs Used to Calculate Mean Quotients	Percentage of Toxicity Test Sampling Locations Above Threshold	*Chironomus*				*Hyalella*				All Four Endpoints Combined (%)	
		Mortality (%)		Biomass (%)		Mortality (%)		Biomass (%)			
		False Positive	False Negative	False Positive	False Negative	False Positive	False Negative	False Positive	False Negative	False Positive	False Negative
PEC	10	6.3	46	5.4	54	6.2	24	6.8	27	5.1	53
PEL	15	11	46	9.7	49	11	29	12	27	9.3	47
RSET_SL2	9	6.3	63	5	63	6.2	47	6.1	40	4.7	61
ERM	11	8.2	54	7.8	63	8.3	41	9	47	7.4	61
CSL	6	3.7	63	3.1	69	4.3	59	4.3	53	3.1	69

CSL – cleanup screening level
ERM – effects range – median
PEC – probable effects concentration
PEL – probable effects level
RSET – Regional Sediment Evaluation Team
SL2 – screening level 2
SQG – sediment quality guideline

Table 3-2. False Positive and Negative Rates Associated with Each High SQG Quotient Based on a Comparison to Low Reference Thresholds

SQGs Used to Calculate Mean Quotients	Percentage of Toxicity Test Sampling Locations Above Threshold	*Chironomus*				*Hyalella*				All Four Endpoints Combined (%)	
		Mortality (%)		Biomass (%)		Mortality (%)		Biomass (%)			
		False Positive	False Negative	False Positive	False Negative	False Positive	False Negative	False Positive	False Negative	False Positive	False Negative
PEC	10	5.5	57	4.1	58	5.9	30	6	48	4.3	67
PEL	15	9.8	51	8.2	52	11	35	11	52	8.6	62
RSET_SL2	9	5.1	65	3.7	65	5.9	50	6	63	3.9	72
ERM	11	7.8	65	6.9	67	8.4	50	8.6	63	7.3	73
CSL	6	3.1	70	3.3	77	4.4	65	4.5	74	3.4	82

CSL – cleanup screening level
ERM – effects range – median
PEC – probable effects concentration
PEL – probable effects level
RSET – Regional Sediment Evaluation Team
SL2 – screening level 2
SQG – sediment quality guideline

Because none of the mean quotients using the thresholds of 1 or 0.7 met the acceptability criterion of ≤ 20% false negative rates, use of lower mean quotient thresholds was evaluated. Similar to the approach used in other studies (Long et al. 2006), the mean quotient thresholds were adjusted in an attempt to achieve acceptable error rates in the comparison to toxicity test results based on the high and low reference thresholds. Because the acceptability criteria for both false negative and false positive rates could not be met, even with adjustments to the quotient threshold, the high SQG mean quotients were not used in the benthic risk characterization.

4.0 RELIABILITY ANALYSIS OF ESBS

ESBs are calculated for PAH mixtures, nonionic organic compounds, and selected pesticides. The predictive power of the ESB for PAH mixtures was assessed by evaluating the false positive and false negative rates based on a comparison of the empirical toxicity test results and the ESB predictions of toxicity. Table 4-1 presents the comparison between the ESB results for PAH mixtures and the adverse effects measured in the 293 toxicity tests based on the high and low reference thresholds. The ESB for PAH mixtures predicted that 18% of the sampling locations had adverse effects on benthic invertebrates. Compared to the toxicity test results for both the high and low reference thresholds, all endpoints had low false positive rates (12 to 16%) and high false negative rates (40 to 63%). Combining the toxicity test results across all four endpoints (adverse effects in any endpoint is counted as adverse effect for that sample) had only slight effects on the false positive and negative rates. Similar to the mean quotient results, the thresholds were adjusted in an attempt to achieve acceptable error rates in the comparison to toxicity test results based on the high and low reference thresholds. In both comparisons the acceptability criterion of ≤ 20% false positive and false negative rates could not be met; hence, the ESB for PAH mixtures was not carried forward in the evaluation of risks to the benthic community.

Table 4-1. False Negative and False Positive Rates Associated with ESB PAH Based on High Reference Thresholds

		Chironomus		***Hyalella***		**Four Endpoints Combined**
ESB Group	**Parameter**	**Mortality**	**Biomass**	**Mortality**	**Biomass**	
PAH mixtures	**High toxicity thresholds**					
	% False negative	46	43	41	40	44
	% False positive	14	12	15	15	12
	Low toxicity thresholds					
	% False negative	46	54	40	63	63
	% False positive	12	12	15	16	13
gamma-HCH	**High toxicity thresholds**					
	% False negative	75	83	67	50	86
	% False positive	22	23	22	21	24
	Low toxicity thresholds					
	% False negative	83	89	67	80	91
	% False positive	23	24	22	23	26

Table 4-1. False Negative and False Positive Rates Associated with ESB PAH Based on High Reference Thresholds

ESB Group	Parameter	*Chironomus* Mortality	*Chironomus* Biomass	*Hyalella* Mortality	*Hyalella* Biomass	Four Endpoints Combined
Sum of PAHs and narcotic nonionic organics	**High toxicity thresholds**					
	% False negative	46	43	41	40	44
	% False positive	14	12	15	15	12
	Low toxicity thresholds					
	% False negative	46	54	40	63	63
	% False positive	12	12	15	16	13

ESB – equilibrium partitioning sediment benchmark
HCH – hexachlorocyclohexane
PAH – polycyclic aromatic hydrocarbon

For the nonionic organic chemicals, a reliability analysis was only performed for gamma-HCH and the sum of PAHs and narcotic nonionic organics because all the other nonionic organic chemicals had low exceedance frequencies (< 5%), were measured only in a limited number of sediment samples, or both. The predictive power of the ESB for gamma-HCH and the sum of PAHs and narcotic nonionic organics was assessed by evaluating the false positive and false negative rates based on a comparison of the toxicity test and the ESB results. Table 4-1 presents the comparison of the ESB for gamma-HCH and the adverse effects in the 293 toxicity tests based on the high and low reference thresholds. The ESB for gamma-HCH predicted that 22% of the sampling locations had adverse effects on benthic invertebrates. Compared with the toxicity test results based on both the high and low reference thresholds, all endpoints had relatively low false positive rates (21 to 24%) and high false negative rates (67 to 91%). Pooling the toxicity test results across all four endpoints increased the rates slightly. Similar to the reliability analysis of the ESB for PAH mixtures, the thresholds were adjusted in an attempt to achieve acceptable error rates in the comparison to toxicity test results based on the high and low toxicity thresholds. In both comparisons, the acceptability criterion of ≤ 20% false positive and false negative rates could not be met.

Table 4-1 presents the comparison between the ESB for the sum of PAHs and narcotic nonionic organics, collectively, and the adverse effects in the 293 toxicity tests based on the high and low toxicity thresholds. The ESB for the sum of PAHs and narcotic nonionic organics predicted that 18% of the sampling locations had adverse effects to benthic invertebrates. A comparison with the toxicity test results based on both the high and low reference thresholds yielded the same false positive and false negative rates as the ESB PAH mixture. Similar to the reliability analysis of the ESB for PAH mixtures, the thresholds were adjusted in an attempt to achieve acceptable error rates in the comparison to toxicity test results based on the high and

low toxicity thresholds. In both comparisons, the acceptability criterion of ≤ 20% false positive and false negative rates could not be met. Because the ESB for gamma-HCH and sum of PAHs and narcotic nonionic organics did not meet the acceptability criteria for the false positive and false negative rates, these SQGs were not carried forward in the evaluation of risks to the benthic community.

5.0 RELIABILITY ANALYSIS OF LOW SQGS

The ability of the sets of low SQGs to predict the absence of toxicity was evaluated based on a comparison of the SQG exceedances and the low toxicity reference thresholds (Section 6.1.1.1 of the BERA and Attachment 6). Low SQGs, described in Table 5-1, included threshold effects concentration (TEC), threshold effects level (TEL), sediment quality standard (SQS), lowest apparent effects threshold (LAET), effects range – low (ERL), and screening level 1 (SL1) thresholds. Because the narrative intent of the low SQGs is to identify the concentrations of COPCs or COPC mixtures below which adverse effects on benthic invertebrates would be infrequently observed, the analysis of the low SQGs focused on false negative rates for each set of low SQGs. The false negative rate is the number of toxicity test samples that were predicted to be non-toxic by the SQGs, but were toxic based on the toxicity test results, divided by the sum of correctly predicted toxic locations and the incorrectly predicted non-toxic locations (i.e., all those samples that were actually toxic). Hence, this error rate identifies the proportion of samples that was erroneously deemed as being clean. Similar to the reliability analysis of the high SQGs, an acceptability criterion of ≤20% was selected for false negative rates.

Table 5-1. Generic SQGs and Their Derivation and Narrative Intent

SQG	Toxicity Threshold	Narrative Intent	Source
TEC	Low	The consensus-based TECs were intended to identify the concentration of sediment-associated contaminants below which adverse effects on sediment-dwelling organisms are not expected to occur. The TECs were derived as geometric means of SQGs (including TELs and ERLs) in the literature with similar narrative intent. Thus, the SQGs are derived from a combination of freshwater and marine toxicity tests.	Macdonald et al. (2000)
TEL	Low	The TELs were intended to estimate the concentrations of chemicals below which adverse biological effects only rarely occurred. The TELs were derived by calculating the geometric mean of the 15th percentile of the effect dataset and the 50th percentile of the no-effect dataset. These SQGs are derived based on a national biological effect database that includes freshwater toxicity tests and changes in freshwater benthic community structures.	Smith et al. (1996)

Table 5-1. Generic SQGs and Their Derivation and Narrative Intent

SQG	Toxicity Threshold	Narrative Intent	Source
SL1	Low	The SL1s correspond to concentrations below which adverse effects on benthic organisms would not be expected. The SL1s were derived using the FPM and freshwater toxicity tests with both mortality and growth endpoints. These SQGs are derived based on a regional (Western Washington and Oregon) biological effect database.	RSET (2006)
ERL	Low	The concentrations below the ERLs represent minimal-effects thresholds that were intended to estimate conditions below which effects would be rarely observed. The ERLs were derived as the 10th percentile of a database composed of multiple studies, species, and effects endpoints (predominantly mortality). These SQGs are derived based on a national biological effect database that includes other sediment SQGs, marine toxicity tests, and benthic field studies.	Long et al. (1995)
SQS	Low	The SQS corresponds to a concentration that will result in no adverse effects, including no acute or chronic adverse effects on biological resources and no significant health risk to humans. SQS are generally based on the LAETs (see LAET). These SQGs are derived based on a biological effect database from Puget Sound, WA.	Ecology (1995)
LAET	Low	AET is the sediment contaminant concentration above which statistically significant ($p \leq 0.05$) adverse biological effects (relative to appropriate reference conditions) would always be expected. LAET is the lowest of four AETs derived from four marine toxicity tests (amphipod mortality, echinoderm and oyster abnormality, and Microtox) and benthic community data. The biological effect database is from Puget Sound, WA.	Gries and Waldow (1996)

AET – apparent effects threshold
Ecology – Washington State Department of Ecology
ERL – effects range – low
FPM – floating percentile model
LAET – lowest apparent effects threshold
SL1 – screening level 1
SQG—sediment quality guideline
SQS – sediment quality standard
TEC – threshold effects concentration
TEL – threshold effects level
RSET – Regional Sediment Evaluation Team

Table 5-2 presents the false negative rates associated with each set of low SQGs. Three sets of low SQGs (TECs, ERLs, and TELs) met the acceptability criterion of a ≤ 20% false negative rate. Maps 5-1, 5-2, and 5-3 present the stations in the Study Area where chemical concentrations did not exceed any of the three sets of low SQGs.

Table 5-2. Number of Stations Predicted to be Non-Toxic Based on the Sets of Low SQGs

SQG Set	Percentage of Toxicity Test Stations Below Threshold	False Negative Rates Relative to Empirical Toxicity Test Results			
		Chironomus		*Hyalella*	
		Mortality (%)	Biomass (%)	Mortality (%)	Biomass (%)
TEC	11	5.4	0	0	0

Table 5-2. Number of Stations Predicted to be Non-Toxic Based on the Sets of Low SQGs

SQG Set	Percentage of Toxicity Test Stations Below Threshold	False Negative Rates Relative to Empirical Toxicity Test Results			
		Chironomus		***Hyalella***	
		Mortality (%)	**Biomass (%)**	**Mortality (%)**	**Biomass (%)**
TEL	3	0	0	0	0
RSET SL1	42	24	15	20	22
ERL	6	2.7	0	0	0
SQS	56	30	27	25	48
LAET	55	32	25	20	37

ERL – effects range – low
LAET – lowest adverse effect threshold
RSET – Regional Sediment Evaluation Team
SL1 – screening level 1
SQG – sediment quality guideline
SQS – sediment quality standards
TEC – threshold effects concentration
TEL – threshold effects level

6.0 REFERENCES

Ecology. 1995. Sediment management standards. WAC 173-204. Washington State Department of Ecology, Toxics Cleanup Program, Sediment Management Unit, Olympia, WA.

Gries TH, Waldow KH. 1996. Progress re-evaluating Puget Sound apparent effects thresholds (AETs). Volume I: 1994 amphipod and echinoderm larval AETs. Draft report. Prepared for Puget Sound Dredged Disposal Analysis (PSDDA) agencies. Washington Department of Ecology, Olympia, WA.

Long ER, MacDonald DD, Smith SL, Calder FD. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. Environ Manage 19:81-97.

MacDonald DD, Ingersoll CG, Berger TA. 2000. Development and evaluation of consensus-based sediment quality guidelines for freshwater ecosystems. Arch Environ Contam Toxicol 39(5):20-31.

RSET. 2006. Regional sediment evaluation framework for the Pacific Northwest. Interim final. September 2006. Regional Sediment Evaluation Team (US Army Corps of Engineers - Seattle District, Portland District, Walla Walla District, and Northwestern Division; US Environmental Protection Agency Region 10; Washington Department of Ecology; Washington Department of Natural Resources; Oregon Department of Environmental Quality; Idaho Department of Environmental Quality; National Marine Fisheries Service; US Fish and Wildlife Service).

Smith SL, MacDonald DD, Keenleyside KA, Ingersoll CG, Field LJ. 1996. A preliminary evaluation of sediment quality assessment values for freshwater ecosystems. J Great Lakes Res 22:624-638.