SETAC PRESS

Environmental Toxicology and Chemistry, Vol. 15, No. 7, pp. 1232–1241, 1996
Printed in the USA
0730-7268/96 $6.00 + .00

Hazard/Risk Assessment

TOXICOLOGICAL BENCHMARKS FOR SCREENING CONTAMINANTS OF POTENTIAL CONCERN FOR EFFECTS ON FRESHWATER BIOTA

G.W. SUTER II
Environmental Sciences Division, Oak Ridge National Laboratory, P.O. Box 2008, MS 6038, Oak Ridge, Tennessee 37831, USA

(*Received* 25 *May* 1995; *Accepted* 2 *November* 1995)

Abstract—An important early step in the assessment of ecological risks at contaminated sites is the screening of chemicals detected on the site to identify those that constitute a potential risk. Part of this screening process is the comparison of measured ambient concentrations to concentrations that are believed to be nonhazardous, termed ''benchmarks.'' This article discusses 13 methods by which benchmarks may be derived for aquatic biota and presents benchmarks for 105 chemicals. It then compares them with respect to their sensitivity, availability, magnitude relative to background concentrations, and conceptual bases. Although some individual values can be shown to be too high to be protective and others are too low to be useful for screening, none of the approaches to benchmark derivation can be rejected without further definition of what constitutes adequate protection. The most appropriate screening strategy is to use multiple benchmark values along with background concentrations, knowledge of waste composition, and physicochemical properties to identify contaminants of potential concern.

Keywords—Benchmarks Ecological risk assessment Screening Criteria

INTRODUCTION

An important early step in the assessment of ecological risks posed by a contaminated site is the screening of chemicals. In many cases concentrations will be reported for more than 100 chemicals, most of which will be reported as undetected at some defined limit of detection. The assessor must decide which of the detected chemicals constitute an ecotoxicological hazard and, because limits of detection may be too high, which of the undetected chemicals may pose a hazard. This screening is done using one or more of the following criteria. If the concentrations are not greater than background concentrations, the chemical may be ignored. If the chemical was not detected and the analytical method was judged to be acceptable, the chemical may be ignored. If the wastes deposited at the site are well specified, chemicals that are not constituents of the waste may be ignored. If the chemical concentration is below concentrations that constitute an ecotoxicological hazard, they can be ignored.

Use of any of these criteria depends on prior agreement among the parties involved in making the risk-management decisions. The first criterion requires definition of the background concentration in a satisfactory manner, agreement on a definition of exceedence, and agreement that the chemical does not exist in a more toxic form on the site than in background sites [1]. The second depends on agreement about the adequacy of the detection limits provided by the proposed analytical methods. For example, some U.S. Environmental Protection Agency (EPA) regional offices use the Contract Laboratory Program Practical Quantification Limits to screen contaminants, even when they are higher than toxic concentrations. The third requires having good records of the wastes deposited on the sites, reasonable assurance that no unrecorded releases occurred, and well-characterized wastes. The fourth criterion depends on definition of chemical concentrations in ambient media that are reliably protective but are not so low as to retain all detected chemicals.

This article is concerned with the last screening criterion. It discusses alternative approaches for calculating ecotoxicological screening benchmarks, presents benchmarks values for some chemicals, and compares their relative sensitivity. This article is limited to benchmarks for screening aqueous chemicals for their hazard to aquatic life. Benchmarks for sediments, soil, and wildlife food and water are presented in Oak Ridge National Laboratory (ORNL) reports [2–5]. All of these benchmarks are regularly updated, and the current versions with supporting documentation can be found on the World Wide Web at http://www.hsrd.ornl.gov/ecorisk/ecorisk.html. In addition, screening benchmarks for some chemicals and media are available from some EPA regional offices and other regulatory agencies. Because the acceptability of screening benchmarks depends on the policies and judgements of the various regulators and responsible parties involved in risk-management decisions at particular sites, this article cannot state which benchmarks should be used. Rather, it attempts to make clear the strengths and weaknesses of alternative methods for deriving benchmarks.

The need to consider alternative benchmarks arises because there are no national benchmarks and little experience or consensus on what constitutes good screening benchmarks. The only values consistently used to screen aqueous contaminants in the United States are the U.S. National Ambient Water Quality Criteria for Protection of Aquatic Life (NAWQC), but they were not designed for that purpose. They are regulatory values that are intended to protect most aquatic species most of the time with reasonable confidence [6]. Because screening benchmarks are intended to minimize the likelihood of screening out a chemical that is hazardous, greater conservatism is warranted. More importantly, NAWQC are available for only a small proportion of chemicals.

This compilation is limited to chemicals that have been detected on the U.S. Department of Energy's Oak Ridge Reservation (ORR) and to benchmarks derived from studies of toxic effects on freshwater organisms. The list of chemicals includes 45 metals and 56 industrial organic chemicals but only four pesticides (chlordane, dichlorodiphenyltrichloroethane [DDT], heptachlor, and lindane).

METHODS FOR DERIVING BENCHMARKS

Types of benchmarks

The simplest screening benchmarks are toxicity test end points. A test end point is a statistically derived numeric summary of the results of a toxicity test. Test end points can be calculated in two ways. First, a level of effect can be estimated by fitting a function such as the probit or logit to the concentration–response data to derive a concentration–response model. Then, by inverse regression a concentration can be estimated that causes a particular level of effect, such as the median lethal concentration (LC50). Second, hypothesis-testing statistics can be used to determine whether each of the tested concentrations caused an effect that was significantly different statistically from the controls. The lowest concentration causing such an effect is termed the ''lowest-observed-effect concentration'' (LOEC); the highest concentration for which there were no such effects is termed the ''no-observed-effect concentration'' (NOEC). The geometric mean of the LOEC and NOEC is termed the ''chronic value'' (CV). Since the NOEC and LOEC are tested concentrations, the benchmarks derived from these values are funtions of the test regime chosen by the toxicologist who designed the test.

Another important distinction is between response-specific and integrative end points. Conventionally, NOECs and LOECs are calculated for each response parameter, and the results for the most statistically sensitive parameter are reported. Because effects on populations and ecosystems are a result of the integrated effects of the toxicant on all life stages, it is more sensible to integrate the responses in the test when calculating the test end point. Integrative end points may be simple arithmetic combinations of effects such as the proportional mortality across all tested life stages or population parameters derived from simple models such as the intrinsic rate of natural increase, *r*.

Benchmarks may be combinations of multiple test end points. An example is the chronic NAWQC, which are derived from at least eight LC50s and three CVs [6]. One common approach to combining test end points is to estimate percentiles or other parameters of the distribution of the end points [6–8]. Finally, benchmarks may be derived by using mathematical models to simulate an assessment end point, a specific environmental characteristic that is valued and is at risk due to the contamination or disturbance that is being assessed [9]. For example, in this study we present concentrations estimated to correspond to a 25% reduction in recruit abundance for largemouth bass (*Micropterus salmoides*) because production of fish, particularly game fish, is an assessment end point for ORR ecological risk assessments [10].

Conventional aquatic benchmarks, which are based on regulatory criteria or standard test end points used to derive criteria, are listed in Table 1. Unconventional aquatic benchmarks, which are based on levels of effects on integrative end points, are listed in Table 2.

Water quality criteria

Because the NAWQC are regulatory standards in the United States, regulators in the United States are likely to require that any chemicals occurring at concentrations that exceed NAWQC be retained in the risk assessment. The acute NAWQC are calculated by the EPA as half the final acute value (FAV), which is the fifth percentile of the distribution of 48- to 96-h LC50 values or equivalent median effective concentration (EC50) values for each criterion chemical [6]. The acute NAWQC are intended to correspond to concentrations that would cause less than 50% mortality in 5% of exposed populations in a relatively brief exposure. They may be used as a reasonable upper screening benchmark because waste site assessments are concerned with sublethal effects and largely with continuous exposures rather than the lethal effects and episodic exposures to which the acute NAWQC are applied. The chronic NAWQC are the FAVs divided by the final acute–chronic ratio (FACR), which is the geometric mean of quotients of at least three LC50/CV ratios from tests of different families of aquatic organisms [6]. It is intended to prevent significant toxic effects in most chronic exposures. The NAWQC are listed in Table 1.

Some chronic NAWQC are based on protection of humans or other piscivorous organisms rather than protection of aquatic organisms (i.e., final residue values). Those criteria are not included here because screening for risks to wildlife or humans is performed by other methods. However, if sufficient data were available to calculate a final chronic value (FCV) for those chemicals, it is presented in place of the chronic NAWQC in Table 1, and its derivation is noted.

For particular chemicals the screening benchmark could be lower than the chronic NAWQC for any one of the following reasons. First, the chronic NAWQC are based on a threshold for statistical significance rather than biological significance. In many chronic tests the LOEC corresponds to greater than 50% effect on a response parameter [11,12]. Second, not all important responses are included in the subchronic toxicity tests that are used to calculate many chronic NAWQC. In particular, effects on fecundity, which is the most sensitive response parameter on average in fish toxicity tests [12], are often not included. Third, the chronic NAWQC are based on the most statistically sensitive of the measured response parameters in each chronic or subchronic test. Therefore, cumulative effects over the life cycle of fish and invertebrates are not considered. Finally, many of the NAWQC have not been revised since 1980, so they do not incorporate recent data that are included in the calculation of other benchmarks. These concerns are supported by the recent finding that nickel concentrations on the ORR that are below chronic NAWQC are nonetheless toxic to daphnids [13].

Tier II values

If NAWQC were not available for a chemical, a slight variation of the Tier II method described in the EPA's ''Water Quality Guidance for the Great Lakes System and Correction: Proposed Rules'' was applied [14]. Tier II values were developed so that aquatic life criteria could be established with fewer data than are required for the NAWQC. The Tier II values presented in this report are concentrations that would be expected to be higher than NAWQC in no more than 20% of cases if sufficient test data were obtained to calculate the NAWQC.

The Tier II values equivalent to the FAV and FCV are the secondary acute values (SAVs) and the secondary chronic values (SCVs), respectively. The sources of data for the Tier II values and the procedure and factors used to calculate the SAVs and SCVs are presented by Suter and Mabrey [15]. The methods differ from those in the Great Lakes guidance [14] in two respects. First, the Great Lakes SAVs require an LC50 for a daphnid, but that requirement would severely restrict the number of benchmarks that could be calculated. The EPA has provided factors for calculating SAVs when no daphnid LC50s are available, and those factors are used herein (C.E. Stephan, personal communication). Second, the calculation of SAVs for the Great Lakes requires high-quality standard LC50 and EC50 values.

Table 1. Summary of conventional benchmarks for priority contaminants in fresh water (all values are μg/L)

Chemical	NAWQ criteria		Tier II values		Lowest chronic value			
	Acute	Chronic	Secondary acute value	Secondary chronic value	Fish	Daphnids	Nondaphnid invertebrates	Aquatic plants
Aluminum	750	87			3,288	1,900		460
Ammonia	pH and temperature dependent				1.7	630		2,400
Antimony			985	104	1,600	5,400[c]		610
Arsenic III	360	190			2,962	914.1		2,320
Arsenic V			170	8.11	891.6	*450		48
Barium			69.1	3.8		5,800[c]		
Beryllium			271	5.09	*57	5.3		100,000
Boron			11,000	547		8,830		
Cadmium	3.9+	1.1+			1.7	0.15		2
Calcium						116,000[c]		
Chromium	1,700+	210+			68.6	<44		397
Chromium VI	16	11			73.2	6.132		2
Cobalt			195	3.06	290	5.1		
Copper	18+	12+			3.8	0.23	6.066	1
Cyanide	22	5.2			7.8		18.33	30
Fluorine			19,200	1,180	*8,784	4,400		
Iron		1,000			1,300[c]	158		
Lead	82+	3.2+			18.88	12.26	25.46	500
Magnesium						82,000[c]		
Manganese			1,470	80.3	1,770	<1,100		
Mercury, inorganic or total	2.4			1.30[b]	<0.23	0.96		5
Mercury, methyl			0.115	0.003	0.52	<0.04		0.8–4.0
Molybdenum			10,100	239		880		
Nickel	1,400+	160+			<35	<5	128.4	5
Potassium						53,000[c]		
Selenium	20	5			88.32	91.65		100
Silver	4.1+			0.36[d]	0.12	2.6		30
Sodium						680,000[c]		
Strontium			6,100	620		42,000[c]		
Thallium			164	18.0	57	130		100
Tin			2,680	73.7		350[c]		
Uranium			33.5	1.87	*142			
Vanadium			284	19.1	80	>940		
Zinc	120+	110+			36.41	46.73	>5,243	30
Zirconium			982	54.9	*548			
Organics								
Acenaphthene	80[e]	23[e]			74	*6,646	227	520
Acetone			200,000	11,200	*507,640	*3,114,182		
Anthracene			0.024	0.0013	*0.09	<2.1		
Benzene			815	45.5	8,250	>98,000		525,000
Benzidene			69.1	3.86	*134			
Benzo[*a*]anthracene			0.49	0.027		*0.65		
Benzo[*a*]pyrene			0.24	0.014		*0.30		
Benzoic acid			743	41.6	*12,976			
Benzyl alcohol			1,050	58	*589			
BHC (lindane)	2.0	0.08			14.6	14.5	3.3	500
BHC (other)			43.6	2.44		*95		
Bis(2-ethylhexl)phthalate			286	32.2	8.4	<3		
2-Butanone			372,000	20,800	*282,170	*1,394,927		
Carbon disulfide			159	8.89	*9,538	*244		
Carbon tetrachloride			4,090	229	1,970[c]	5,580[c]		
Chlordane	2.4	0.17[a]			1.6	16	1.09	
Chlorobenzene			2,270	127	*1,203	*15,042		224,000
Chloroform			3,360	188	1,240	*4,483		
DDD p,p′			0.18	0.010	*1.69			
DDT	1.1			0.04[b]	0.73[c]	*0.016		0.3
Decane			878	49		*7,874		
Di-*n*-butyl phthalate			234	32.7	717[d]	697		
Dibenzofuran			365	20.4		*1,003		
1,1-Dichloroethene			834	46.6	*14,680			
1,2-Dichloroethane			13,500	1,100	41,364	15,200		
1,1-Dichloroethene			3,520	196	>2,800	*4,720		>798,000
1,2-Dichloroethenes			558	31.2	*9,538			
1,3-Dichloropropene			459	25.6	244	*805		4,950
Diethyl phthalate			3,950	220				85,600
Di-*n*-octyl phthalate					3,822	708		
Ethyl benzene			5,269	294	>440	*12,922		>438,000

Table 1. Continued

Chemical	NAWQ criteria		Tier II values		Lowest chronic value			
	Acute	Chronic	Secondary acute value	Secondary chronic value	Fish	Daphnids	Nondaphnid invertebrates	Aquatic plants
Fluoranthene	33.6[e]	6.16[e]			30	15		54,500
Heptachlor	0.52			0.029[b]	1.26	*3.18		26.7
Hexane			3,390	189	*65,712			
2-Hexanone			1,770	98.8	*32,783			
1-Methylnaphthalene			37.2	2.08	*526			
4-Methyl-2-pentanone			2,100	164	77,400			
2-Methylphenol			1,290	72.2	*489	*1,316		
Methylene chloride			25,600	2,240	108,000	*42,667		
Naphthalene			353	23.4	620	*1,163		33,000
4-Nitrophenol			1,580	163	*481	7,100		4,190
N-Nitrosodiphenylamine			439	24.5	*332	*1,042		
3-Octanone			6,060	338	*5,449	*110,147		
PCBs, total	2.0			0.19[b]	0.2	2.1[c]	0.8	0.1
Aroclor® 1221			4.83	0.27	*60			4,400
Aroclor® 1232			9.01	0.50	*124			
Aroclor® 1242			0.75	0.06	9.0		4.9	300
Aroclor® 1248			0.16	0.01	0.2	4.3[c]	3.3	
Aroclor® 1254			0.21	0.02	1.0	2.1[c]	0.8	0.1
Aroclor® 1260			187	10.5	<1.3			
1-Pentanol			6,170	344	*30,493			
Phenanthrene			37.1	3.23		200		
Phenol			2,010	117	<200	*2,005		20,000
2-Propanol			41.4	2.31	*590			
1,1,2,2-Tetrachloroethane			2,149	418	2,400	9,900		136,000
Tetrachloroethene			998	125	840	750		>816,000
Toluene			2,383	133	*1,269	*25,229		245,000
1,1,1-Trichloroethane			617	62.1	*3,493	1,770[c]		>669,000
1,1,2-Trichloroethane			6,940	1,400	9,400	18,400		
Trichloroethene			3,288	351	14,867	*7,257		
Vinyl acetate			372	20.8	*810			
Vinyl chloride			1,570	87.8	*28,879			
Xylene			1,540	86.2	*62,308			

* Numbers preceded by * are estimates. Methods of estimation are described in the text.
+ Hardness-dependent criterion normalized to 100 mg/L.
[a] The chronic NAWQC for chlordane (0.0043 μg/L) is based on the final residue values. The FCV is used as a benchmark to protect aquatic life.
[b] The chronic NAWQC for DDT (0.001 μg/L), inorganic mercury (0.012 μg/L), total PCBs (0.014 μg/L), and heptachlor (0.0038 μg/l) are based on the final residue values; for benchmarks to protect aquatic life, we calculated SCVs.
[c] Benchmarks based on tests that are not standard but are judged to be of good quality.
[d] Based on three acute/chronic ratios which were judged by the EPA to be too uncertain for derivation of a chronic NAWQC but are judged to be more reliable than the default ratios for calculation of an SCV.
[e] These numbers are FAVs and FCVs calculated by the EPA for use in the derivation of sediment-quality criteria [23, 24].

Only high-quality standard data are used in this document if such values are available for a chemical. However, when no such values are available, nonstandard LC50s or EC50s are used if the deviation from standard methods is not expected to result in a higher end point value. This deviation is justified by the use of the SAVs derived herein for screening purposes as opposed to the SAVs for the Great Lakes, which are intended for regulatory purposes.

Lowest chronic values

Chronic values are used to calculate the chronic NAWQC and are presented in place of chronic criteria by the EPA when chronic criteria cannot be calculated. Except where noted, the CVs for fish and invertebrates meet the EPA's standards for acceptability [6]. Because of the relative lack of data from standard chronic tests for aquatic plants, EPA guidelines are followed in using any algal test of at least 96-h duration and any biologically meaningful response for the plant values [6].

Estimated lowest chronic values

When acute but not chronic toxicity data are available for a chemical, estimated lowest chronic values for fish and invertebrates are potential benchmarks. Estimated chronic values were extrapolated from 96-h LC50s using equations derived by regression analysis [12,16]. The equations are as follows, where LC50 is the lowest species mean 96-h LC50 for fish, EC50 is the lowest 48-h EC50 for daphnids, and CV is the estimated chronic value for the taxon. The log-scaled 95% prediction interval (PI) at the mean is log CV ± the PI value (95% PIs contain 95% of observations vs. 95% confidence intervals, which contain the mean with 95% confidence).

Fish CV for a metallic contaminant:

$$\log CV = 0.73 \log LC50 - 0.70$$
$$PI = 1.2 \quad (1)$$

Fish CV for an organic contaminant:

$$\log CV = 107 \log LC50 - 1.51$$
$$PI = 1.5 \quad (2)$$

Table 2. Summary of alternative benchmarks for priority contaminants in freshwater based on levels of chronic effects (all values are μg/L)

Chemical	Lowest test EC20		Sensitive species test EC20	Population EC25
	Fish	Daphnids		
Aluminum	4,700	540	75	
Antimony	2,310	1,900		79
Arsenic III	2,130	633	55	1,995
Arsenic V	1,500	>932		185
Barium				
Beryllium	*148	3.8		21
Boron		7		
Cadmium	1.8	0.75	0.013[a]	4.3
Calcium				
Chromium III	89		8.44	126
Chromium VI	51	0.5	0.266	316
Cobalt	810	<4.4		3.98
Copper	5	0.205	0.26	8.6
Cyanide	5.3		1.17	11
Fluorine	*5,336	3,706		1,080
Iron		16		
Lead	22		0.35	71
Magnesium				
Manganese	1,270	<1,100		112
Mercury, inorganic	0.87	0.87	0.18	0.32
Mercury, methyl	<0.03	0.87		0.28
Molybdenum		360		
Nickel	62	45	11[a]	215
Potassium				
Selenium	40	25	2.60	
Silver	0.20	<0.56	0.14[a]	0.32
Sodium				
Strontium				
Thallium	81	64		67
Tin				
Uranium	*455			27
Vanadium	41	430		32
Zinc	47		21	80
Zirconium	*2,396			251
Organics				
Acenaphthene	<197			
Acetone	*161,867			23,714
Anthracene	0.35	>8.2		
Benzene	21[c]			229
Benzidene	*158			68
Benzo[*a*]anthracene				
Benzo[*a*]pyrene	>2.99			
Benzoic acid	*7,409			1,259
Benzyl alcohol	*550			375
BHC (lindane)	<1.1	11	0.11	
BHC (other)				
Bis(2-ethylhexl)phthalate	>54	<3		50
2-Butanone	*98,722			17,783
Carbon disulfide	*5,719			1,000
Carbon tetrachloride	65[c]			224
Chlordane	<0.25	12.1	0.50	0.71
Chlorobenzene	1,002			165
Chloroethane				
Chloroform	8,400[c]			562
DDD p,p′	*3.99			0.61
DDT	0.35		0.008	
Decane				
Di-*n*-butyl phthalate	270	500		251
Dibenzofuran				
1,1-Dichloroethane	*8,219			1,585
1,2-Dichloroethane	29,000	<11,000		1,259
1,1-Dichloroethene				447
1,2-Dichloroethenes	*5,719			
1,3-Dichloropropene	*350			40
Diethyl phthalate				1,000
Di-*n*-octyl phthalate	<100	310		1,995
Ethyl benzene				398
Fluoranthene				32

Table 2. Continued

Chemical	Lowest test EC20		Sensitive species test EC20	Population EC25
	Fish	Daphnids		
Heptachlor	0.86		0.004	0.1
Hexane	*28,995			
2-Hexanone	*16,155			1,259
1-Methylnaphthalene	*500			31.62
4-Methyl-2-pentanone				1,585
2-Methylphenol	*470			74
Methylene chloride	410			1,259
Naphthalene	450	>600		1,000
4-Nitrophenols	*464	5,000		60
N-Nitrosodiphenylamine	*339			40
3-Octanone	*3,571			
PCBs, total	0.4	1.2[c]		0.63
Aroclor® 1221	*80			10
Aroclor® 1232	*148			16
Aroclor® 1242	<2.9			1.58
Aroclor® 1248	0.4	2.5[c]		1.26
Aroclor® 1254	0.52	1.2[c]		0.63
Aroclor® 1260	2.1			316
1-Pentanol	*15,200			3,548
Phenanthrene		110		
Phenol	<230			4,467
2-Propanol	*35,381			3,162
1,1,2,2-Tetrachloroethane	1,400	<420		1,585
Tetrachloroethene	500	510		50
Toluene	<26[c]			200
1,,1,1-Trichloroethane	*2,457	1,300[c]		251
1,1,2-Trichloroethane	14,800	13,000		15,849
Trichloroethene	5,758			232
Vinyl acetate	*718			108
Vinyl chloride	*14,520			
Xylene	2,680[c]			

* Numbers preceded by * are estimates. Methods of estimation are described in the text.
[a] Study LC50s were used rather than species mean LC50s so water hardness would correspond to EC20 values.
[c] Benchmarks based on tests that are not standard but are judged to be of high quality.

Daphnid CV for a metallic contaminant:

$$\log CV = 0.96 \log EC50 - 1.08$$
$$PI = 1.56 \quad (3)$$

Daphnid CV for an organic contaminant:

$$\log CV = 1.11 \log EC50 - 1.30$$
$$PI = 1.35 \quad (4)$$

Test EC20s

Another potential benchmark is the test EC20 for fish, which is defined as the highest tested concentration causing less than 20% reduction in the weight of young fish per initial female fish in a life-cycle or partial life-cycle test or the weight of young per egg in an early life-stage test. A similar potential benchmark is the test EC20 for daphnids, which is the highest tested concentration causing less than 20% reduction in the product of growth, fecundity, and survival in a chronic test with a daphnid species. These benchmarks are intended to be indices of population production. They are equivalent to chronic values in that they are simply a summary of the results of chronic toxicity tests, and in most cases the same test supplied the lowest chronic value and the lowest test EC20. However, the test EC20s are based on a level of biological effect rather than a level of statistical significance, and they integrate all stages of the toxicity test rather than treating each response independently. The 20% figure was chosen because it is a little lower than the mean level of effect on individual response parameters observed at CVs, and it is a minimum detectable difference in population characteristics in the field [10,12]. These values are listed in Table 2.

Estimated test EC20s for fish

The estimated values were extrapolated from 96-h LC50 values using equations derived by regression analysis [16]. The equation is as follows, where LC50 is the lowest species mean 96-h LC50 for fish and the EC25 for weight of juveniles per egg is used as an estimate of the test EC20 value. (The 25% value was chosen in a prior modeling program as a convenience [12] and is used here because the difference between 20 and 25% effect is trivial given the uncertainties in these estimates and the steepness of the concentration– response curves.) The log-scaled 95% PI at the mean is log EC25 ± the PI value:

$$\log EC25 = 0.90 \log LC50 - 0.86$$
$$PI = 1.6 \quad (5)$$

These values are listed in Table 2 for those chemicals that have no empirical test EC20 for fish.

Sensitive species test EC20s

The sixth potential benchmark is the EC20, adjusted to approximate the fifth percentile of the species sensitivity distri-

Table 3. Comparisons of alternative screening benchmarks for aquatic life on the basis of the number of chemicals for which each could be calculated (*n*); the percentage of those chemicals for which it was the lowest benchmark; and the median, minimum, and maximum of the ratios of benchmark values to the chronic NAWQC. The sum of percent lowest values is greater than 100 because *n* varies.

		Percent lowest values	Ratio to chronic NAWQC		
	n[a]		Median	Minimum	Maximum
Chronic NAWQC	16	19			
Secondary chronic value	93	78			
Fish chronic value					
Measured	44 (7)	6.1	2.52	0.22	182.50
Estimated	31	0			
Daphnid chronic value					
Measured	43 (6)	4.2	0.56	0.02	181.25
Estimated	28	0			
Nondaphnid invertebrate chronic value	11 (1)	0	6.41	0.5	147.66
Aquatic plant value	37 (4)	0	2.38	0.03	6,250
Fish test EC20					
Measured	38 (10)	13	1.14	0.39	54.02
Estimated	29	0			
Daphnid test EC20	29 (9)	8.1	0.70	0.02	137.50
Sensitive species test EC20	17	71	0.11	0.01	2.94
Population EC25	70	8.7	3.02	0.60	28.73

[a] Numbers in parentheses are the number of additional benchmarks of the type that were derived but are not always useful because they are > or < values.

bution. It is calculated in the same way as the chronic NAWQC, except that the test EC20s are used in place of CVs, and saltwater species were not included. The FAV calculated for each of the criterion chemicals by the EPA was divided by the geometric mean of ratios of LC50s to EC20s. These benchmarks are referred to as sensitive species (SS) test EC20s and are listed in Table 2.

Population EC25s

The last potential benchmark is an estimate of the continuous concentration that would cause a 25% reduction in the recruit abundance of largemouth bass. The method used was described by Barnthouse et al. [17] and is briefly summarized here. The recruit abundance estimates are generated by a matrix model of a reservoir largemouth bass population [18]. The fecundity, hatching success, larval survival, and postlarval survival of the model population are each decremented by a value generated from statistical extrapolation models. For each life stage for which a concentration–response relationship could be calculated, that relationship was adjusted for the relative sensitivity of the test species and the bass. For those life stages with no concentration–response relationship, the relationship was estimated using life-stage-to-life-stage extrapolation models, and the taxonomic adjustment was made. However, if the authors of the study reported that life stage was unaffected, the decrement for that life stage was set to zero. If no chronic test data were available, extrapolations from LC50s to chronic responses of each life stage were performed. Uncertainties in all of these extrapolations were propagated through the models to generate estimates of uncertainty. For each chemical, each available freshwater fish chronic test was used to parameterize a model run. If no chronic test data were available, each available freshwater fish LC50 was used to parameterize a model run. The results are presented in Suter and Mabrey [15]. The geometric mean of all population EC25 estimates for each chemical is reported in Table 2.

COMPARISON OF BENCHMARKS

Benchmarks can be compared on the basis of the number of chemicals for which they can be derived, their sensitivity relative to each other and relative to the NAWQC, the frequency with which they are lower than background concentrations, and their appropriateness as estimates of the threshold for aquatic effects. Note that the term sensitivity is used here to indicate the relative magnitudes of the types of benchmarks, whereas the term conservatism is used to indicate whether the benchmark contains safety factors or other model assumptions that were intended to make it more sensitive than a direct estimate of the threshold for toxic effects. Frequencies of availability of each benchmark, the frequency with which each is the most sensitive benchmark, and their magnitudes relative to the NAWQC are presented in Table 3.

The relative utility of the benchmarks is determined in large part by their availability. At least one benchmark could be derived for 92% of the 105 chemicals reviewed. The NAWQC and the SS test EC20 values are available for relatively few of the chemicals detected on the ORR (15 and 16%, respectively) because they require large data sets. The situation is particularly bad for organic chemicals; only 6% have NAWQC. The Tier II values are available for most of the chemicals (88%) because they each require only a single LC50 or EC50. For the same reason, if estimated values are included, the CVs for fish and daphnids and the test EC20s for fish are available for most chemicals. However, if estimated values are excluded, fish and daphnid CVs are available for only 49 and 47% of chemicals, respectively. This lack of information concerning chronic toxicity is disquieting given that estimates of chronic effects from acute effects are highly imprecise (Eqns. 1 through 4).

Only 19% of the chronic NAWQC for protection of aquatic life are the lowest benchmark for these chemicals. This result was expected, given the characteristics of the criteria discussed previously.

Table 4. Frequency of values of each aquatic screening benchmark exceeded by the total concentrations of metals at background sites, mean of 10 Wisconsin rivers, and one uncontaminated stream on the Oak Ridge Reservation [25, 26]. Ratios are the number of chemicals with background concentrations exceeding the benchmark over the number of chemicals for which benchmark values were available. Hardness-dependent criteria were corrected for site-specific hardness for the Melton Branch comparisons, and a default hardness of 100 mg/L was used for the Wisconsin rivers

Benchmark	Wisconsin rivers	Melton Branch
Chronic NAWQC	1/5	1/4
Secondary chronic value	1/1	3/8
Fish lowest chronic value	1/6	0/6
Daphnid lowest chronic value	2/6	3/12
Nondaphnid invertebrate lowest chronic value	0/3	0/1
Plant values	0/6	1/3
Fish EC20	1/6	0/6
Daphnid EC20	2/4	2/8
Sensitive species EC20	5/6	2/3
Population EC25	1/5	1/4

Secondary chronic values are the lowest benchmark for 78% of the chemicals for which they were calculated. This sensitivity was not surprising given the goal of the method of ensuring with 80% confidence that these values would not exceed the NAWQC. However, the relative sensitivity of the Tier II values declines as the number of acute and chronic test data that were used to calculate them increases, so they are highly sensitive for chemicals that have been least tested but are hardly more sensitive than NAWQC for chemicals that lack only one or two of the test end points required for derivation of NAWQC. Note that these two benchmarks are never compared because Tier II values are derived only when there are no NAWQC.

Fish CVs were available for 42% of the 105 chemicals, daphnid CVs for 41%, nondaphnid invertebrate CVs for 10%, and plants values for 35%. Chronic values are intermediate in their sensitivity. In general, CVs for daphnids are lower than the others. On average, the lowest CVs for fish are more than twice the NAWQC, while the lowest CVs for daphnids are approximately half (Table 3). The sensitivity of daphnids has been documented previously [19]. However, for 16 chemicals the lowest fish CV was lower than the lowest daphnid CV (not including estimated values). Estimated CVs, CVs for nondaphnid invertebrates, and plant values were never the lowest benchmark for this set of chemicals.

The test EC20 values differ from the other potential benchmarks in that they represent an observed and specified effect. All other benchmarks are estimated using models (population EC25s, estimated CVs, and test EC20s), correspond to no particular effect (CVs and NAWQC), or are both estimated and correspond to no particular effect (SCVs). However, test EC20 values are not very sensitive, being the lowest benchmark for only 8.7% of chemicals and averaging three times the NAWQC. However, the test EC20 values are lower than the CVs for the same taxa in 62.5 and 82% of chemicals for fish and daphnids, respectively. This result is expected given that the test EC20s integrate across life stages but the CVs do not.

The SS test EC20 values are quite sensitive. These values are on average 11% of the NAWQC when both could be calculated. They are equivalent to the NAWQC but use integrative chronic test end points.

The population EC25 values differ from the other potential benchmarks in that they represent a specified effect on a field population. Although they are the lowest benchmarks in approximately the same proportion of cases as the fish CVs and EC20s (8.7%), on average they are three times the chronic NAWQC.

A benchmark that frequently suggests that there are hazards to aquatic life from chemicals occurring at background levels would not be useful for screening. There is no high-quality data set of national background water concentrations against which the benchmarks can be compared. Therefore, we compared the benchmarks to aqueous concentrations of six metals at background sites in 10 Wisconsin rivers and of 12 metals in a stream used as a background site for the assessment of a contaminated stream on the ORR (Table 4). These studies used clean sampling and analysis techniques to avoid the inflated background metal values reported in some studies [20]. The benchmarks are compared to total metal concentrations rather than dissolved concentrations because total concentrations are required by the EPA regional offices. The frequencies of exceedence by background

Table 5. Major strengths and weaknesses of the alternative benchmarks

Benchmark	Advantages	Disadvantages
Chronic NAWQC	Always acceptable in the United States Seldom below background	Available for few chemicals Not conservative Not sensitive Based on statistical significance
Secondary chronic value	Available for many chemicals Conservative	Often below background Based on statistical significance
Fish, daphnid, and nondaphnid invertebrate lowest chronic values	Conventional Seldom below background Reasonably sensitive if it includes daphnid	Not conservative Based on statistical significance
Plant values	Covers an ecologically important group Population functional effect Seldom below background	Not conservative Seldom sensitive Inconsistent
Fish and daphnid EC20s	Reasonably sensitive Based on specified effects Seldom below background	Unconventional Not conservative
Sensitive species EC20	Based on specified effect	Unconventional Available for few chemicals Often below background
Population EC25	Available for many chemicals Based on specified population effect	Highly unconventional

of the different benchmarks are quite variable. However, all benchmarks except the nondaphnid invertebrate CVs (of which there are few) were lower than background for some metals. The SCV and the SS test EC20 were lower than background relatively frequently. This result reinforces the need to screen against background as well as against ecotoxicological benchmarks.

A final consideration in evaluating the benchmarks is their quality as representatives of the actual threshold for toxicity to freshwater aquatic life. None of the benchmarks are ideal in that regard. Most are based on thresholds for statistical significance rather than biological significance. All of the estimated CVs and EC20 values and most of the SCVs and population EC25s are based on acute test end points which estimate chronic end points with more than order of magnitude of uncertainty. All but the population EC25 are based entirely on organism-level responses in the laboratory, but the extrapolation models required to estimate the response of a bass population in the field make the total uncertainty in the population EC25s quite large because none are based on centrarchid life-cycle tests [21].

CONCLUSIONS

All of the types of benchmarks considered have advantages and drawbacks (Table 5). Just as there is no consistently most- or least-sensitive toxicity test, none of these benchmarks are consistently too sensitive or inadequately sensitive. The types of benchmarks in Table 1 have all been used or proposed for some purpose by some component of the EPA, but none have been adopted as screening benchmarks, and only the NAWQC are enforced. The benchmarks in Table 2 have no support from the national EPA but are more biologically and ecologically based. This set of benchmarks is presented here to show how the choice of method for calculating benchmarks can influence their sensitivity and utility and to encourage more consideration of alternative methods for deriving benchmarks. It has evolved through the experience gained in performing screening assessments for the ORR and through discussions with regulators. Currently, all of the benchmarks are used for screening contaminants of potential concern for aquatic life on the ORR along with background concentrations, information concerning contaminants released by the U.S. Department of Energy, and evaluation of the abundance and quality of the available analytical data [22]. Before using any of these benchmarks for screening purposes at other sites, concurrence of the relevant regulators should be sought.

None of these benchmarks should be assumed to constitute thresholds for significant effects at individual sites for purposes of estimating ecological baseline risks or defining remedial goals. Risk estimation should be based on weight of evidence, including knowledge of the condition of the aquatic community in the receiving water. Interpretation of laboratory toxicity data should be more intensive than in screening assessments and should include consideration of mode of action, temporal variance in exposure relative to toxicokinetics, chemical speciation, and bioavailability.

Acknowledgement—M.A. Futrell, G.A. Kerchner, and J.B. Mabrey assisted with literature searching and calculation of benchmarks. B.E. Sample, D.S. Jones, and two anonymous reviewers provided helpful comments. This research was sponsored by the Office of Environmental Restoration and Waste Management, U.S. Department of Energy, under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc. Publication 4475, Environmental Sciences Division, Oak Ridge National Laboratory.

REFERENCES

1. **LaGoy, P.K.** and **C.O. Schulz.** 1993. Background sampling: An example of the need for reasonableness in risk assessment. *Risk Analysis* **13**:483–484.
2. **Hull, R.N.** and **G.W. Suter II.** 1994. Toxicological benchmarks for screening potential contaminants of concern for effects on sediment-associated biota: 1994 revision. ES/ER/TM-95/R1. Oak Ridge National Laboratory, Oak Ridge, TN, USA.
3. **Will, M.E.** and **G.W. Suter II.** 1995. Toxicological benchmarks for screening potential contaminants of concern for effects on terrestrial plants: 1995 revision. ES/ER/TM-85/R2. Oak Ridge National Laboratory, Oak Ridge, TN, USA.
4. **Will, M.E.** and **G.W. Suter II.** 1995. Toxicological benchmarks for screening potential contaminants of concern for effects on soil and litter invertebrates and heterotrophic processes: 1995 revision. ES/ER/TM-126/R1. Oak Ridge National Laboratory, Oak Ridge, TN, USA.
5. **Opresko, D.M., B.E. Sample** and **G.W. Suter II.** 1995. Toxicological benchmarks for wildlife: 1995 revision. ES/ER/TM-86/R2. Oak Ridge National Laboratory, Oak Ridge, TN, USA.
6. **Stephan, C.E., et al.** 1985. Guidelines for deriving numeric National Water Quality Criteria for the protection of aquatic organisms and their uses. PB85-227049. U.S. Environmental Protection Agency, Washington, DC.
7. **Van Leeuwen, K.** 1990. Ecotoxicological effects assessment in the Netherlands. *Environ. Manage.* **14**:779–792.
8. **van Straalen, N.M.** and **G.A.J. Denneman.** 1989. Ecological evaluation of soil quality criteria. *Ecotoxicol. Environ. Saf.* **18**:241–245.
9. **Suter, G.W., II.** 1989. Ecological endpoints. In W. Warren-Hicks, B.R. Parkhurst, and J.S.S. Baker, eds., *Ecological Assessment of Hazardous Waste Sites: A Field and Laboratory Reference Document.* EPA 600/3-89/013. Corvallis Environmental Research Laboratory, Corvallis, OR, USA, pp. 2-1–2-28.
10. **Suter, G.W., II, et al.** 1995. Approach and strategy for performing ecological risk assessments for the Department of Energy Oak Ridge Field Office Environmental Restoration Program. ES/ER/TM-33/R2. Environmental Restoration Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA.
11. **Stephan, C.E.** and **J.R. Rogers.** 1985. Advantages of using regression analysis to calculate results of chronic toxicity tests. In R.C. Bahner and D.J. Hansen, eds., *Aquatic Toxicology and Hazard Assessment: Eighth Symposium.* American Society for Testing and Materials, Philadelphia, PA, USA, pp. 328–339.
12. **Suter, G.W., II, et al.** 1987. Endpoints for responses of fish to chronic toxic exposures. *Environ. Toxicol. Chem.* **6**:793–809.
13. **Kszos, L.A., A.J. Stewart** and **P.A. Taylor.** 1992. An evaluation of nickel toxicity to *Ceriodaphnia dubia* and *Daphnia magna* in a contaminated stream and in laboratory toxicity tests. *Environ. Toxicol. Chem.* **11**:1001–1012.
14. **U.S. Environmental Protection Agency.** 1993. Water quality guidance for the Great Lakes system and correction: Proposed rules. *Fed. Reg.* **58**:20802–21047.
15. **Suter, G.W., II** and **J.B. Mabrey.** 1994. Toxicological benchmarks for screening potential contaminants of concern for effects on aquatic biota: 1994 revision. ES/ER/TM-96/R1. Oak Ridge National Laboratory, Oak Ridge, TN, USA.
16. **Suter II, G.W.** 1993. *Ecological Risk Assessment.* Lewis, Boca Raton, FL, USA.
17. **Barnthouse, L.W., G.W. Suter II** and **A.E. Rosen.** 1990. Risks of toxic contaminants to exploited fish populations: Influence of life history, data uncertainty, and exploitation intensity. *Environ. Toxicol. Chem.* **9**:297–311.
18. **Bartell, S.M.** 1990. Ecosystem context for estimating stress-induced reductions in fish populations. *Am. Fish. Soc. Spec. Symp.* **8**:167–182.
19. **Host, G.E., R.R. Regal** and **C.E. Stephan.** 1991. Analysis of acute and chronic data for aquatic life. PB93-154748. National Technical Information Service, Springfield, VA, USA.
20. **Windom, H.L., et al.** 1991. Inadequacy of NASQAN data for assessing metal trends in the nation's rivers. *Environ. Sci. Technol.* **25**:1137–1142.
21. **Barnthouse, L.W., G.W. Suter II** and **A.E. Rosen.** 1988. Inferring population-level significance from individual-level effects: An extrapolation from fisheries science to ecotoxicology. In G.W. Suter II and M. Lewis, eds., *Aquatic Toxicology and Hazard Assessment,*

11th Symposium. American Society for Testing and Materials, Philadelphia, PA, USA, pp. 289– 300.
22. **Suter, G.W., II.** 1995. Guide for performing screening ecological risk assessments at DOE facilities. ES/ER/TM-153. Lockheed Martin Energy Systems, Oak Ridge, TN, USA.
23. **U.S. Environmental Protection Agency.** 1993. Sediment quality criteria for protection of benthic organisms—acenaphthene. EPA-822-R-93-013. Washington, DC.
24. **U.S. Environmental Protection Agency.** 1993. Sediment quality criteria for protection of benthic organisms—fluoranthene. EPA-822-R-93-0012. Washington, DC.
25. **Webb, D.A.** 1992. Background metal concentrations in Wisconsin surface waters. Wisconsin Department of Natural Resources, Madison, WI, USA.
26. **U.S. Department of Energy.** 1995. Remedial investigation report on Waste Area Grouping 5 at Oak Ridge National Laboratory, Oak Ridge, Tennessee. ORNL/ER-284. Office of Environmental Restoration and Waste Management, Oak Ridge, TN, USA.