

DECISION PROCESS FOR IDENTIFICATION OF
ESTUARINE BENTHIC IMPAIRMENTS

Submitted to:

Maryland Department of Natural Resources
Resource Assessment Service
Tidewater Ecosystem Assessments
Annapolis, Maryland

Submitted by:

Roberto J. Llansó
Jon H. Vølstad

Versar, Inc.
Columbia, Maryland

Daniel M. Dauer

Department of Biological Sciences
Old Dominion University
Norfolk, Virginia

November 2003

FOREWORD

This report, *Decision Process for Identification of Estuarine Benthic Impairments*, was prepared by Versar, Inc., at the request of the Maryland Department of Natural Resources, Maryland Department of the Environment, Virginia Department of Environmental Quality, and USEPA Region III under Cooperative Agreement CA-03-24-07-4-30831-3734 between Versar Inc., and the University of Maryland Center for Environmental and Estuarine Studies. Dr. Daniel M. Dauer, of Old Dominion University, contributed to project conceptualization and evaluation. Versar wishes to thank members of the Workgroup, including representatives of the above-mentioned agencies, for their support and for contributing guidance to the project. We are grateful to Ms. Jody Dew, Versar's SAS Programmer, for skillfully managing data and conducting statistical analyses within the very short time frame of the study.

TABLE OF CONTENTS

	Page
FOREWORD	ii
1.0 INTRODUCTION	1
2.0 OBJECTIVES	2
3.0 METHOD DEVELOPMENT AND RESULTS	2
3.1 DATA	2
3.2 REFERENCE FREQUENCY DISTRIBUTIONS.....	3
3.3 SEGMENTS WITH MULTIPLE HABITATS.....	4
3.4 RECOMMENDATION FOR STATISTICAL METHOD	4
3.5 EVALUATION OF DATA SETS AND TYPE I AND II ERRORS.....	4
3.6 MINIMUM REQUIRED SAMPLE SIZE	5
3.7 FINAL ASSESSMENT	5
4.0 RECOMMENDATIONS	6
5.0 REFERENCES	6

1.0 INTRODUCTION

To meet the requirements of the Clean Water Act, the States of Maryland and Virginia are in the process of developing biological criteria for evaluating estuarine waters. The Chesapeake Bay Benthic Index of Biotic Integrity (B-IBI) is the basis for these biological criteria. As an initial step, a Workgroup developed an interim framework for the application of the B-IBI to the States' water quality inventory report (305b report) and list of impaired waters (303d list). The Workgroup included representatives of the Maryland Department of the Environment (MDE), Maryland Department of Natural Resources (MDNR), Virginia Department of Environmental Quality (DEQ), USEPA Region III Office of Watersheds, Old Dominion University Department of Biological Sciences, and Versar, Inc.

Two alternative methods were initially proposed for 303(d) impairment decisions based on the B-IBI: (1) a weighted (stratified) mean approach, and (2) comparisons of cumulative frequency distributions and proportions. In the first method, the mean benthic condition of a Chesapeake Bay segment (as determined by the stratified mean B-IBI value) would be compared against the mean of the reference condition (based on the original index development reference sites) using a t-test. Because a segment may be represented by more than one habitat for which reference conditions may differ, weights are applied to combine means for each habitat type (strata) (Cochran 1977). In the second method, the cumulative frequency distribution of B-IBI values for a segment would be compared to the cumulative distribution of B-IBI values for the reference condition, and a test conducted to determine whether the percentage of sites with low B-IBI scores (below a threshold) significantly exceeds the percentage of sites with low scores in the reference distribution, and whether the percent exceedance is above a specified threshold (e.g., 25%).

Each of the above methods presents advantages and disadvantages that can only be evaluated statistically. However, the weighted mean approach is only applicable if each habitat has two or more samples, while the cumulative frequency distribution method can be applied only when the reference distributions for the seven Chesapeake Bay habitats are similar and the data can be pooled into one reference distribution. Examination of the B-IBI reference distributions for the seven habitats indicated significant differences among habitats and small sample sizes. Based on this result, an alternative non-parametric frequency distribution approach that does not require the pooling of the reference data was applied. The method of choice, a stratified Wilcoxon rank sum test, treats the habitats within a segment as strata, and tests for significant shifts in the distribution of B-IBI data between the assessment and the reference B-IBI values. We describe below the method and present each of the analytical steps leading to the assessment and identification of Chesapeake Bay segments with estuarine benthic impairments.

This report is organized into five sections: introduction, objectives, method development and results, recommendations, and references. The method development section is further divided into seven subsections. The first subsection describes the data used in this project, and subsequent subsections follow closely the original SOW tasks, with modifications to account for the new method:

1. Data.
2. Examine reference frequency distributions for similarity.

3. Determine the number of Bay segments having multiple habitats.
4. Provide recommendations for the frequency distribution methodology based on examination of reference curves.
5. Evaluate the proposed approach using actual data sets, with consideration of any differences in results, ease of evaluation, and balance of Type I and Type II errors.
6. Evaluate the power of the proposed approach with respect to small sample sizes and provide recommendation for a minimum required sample size.
7. Produce a final assessment of Chesapeake Bay segments using the options agreed upon by the Workgroup.

2.0 OBJECTIVES

1. To develop a methodology for the assessment of Chesapeake Bay benthic community status for 2004 303(d) impairment decisions.
2. To produce an assessment of aquatic life support for Chesapeake Bay segments.

3.0 METHOD DEVELOPMENT AND RESULTS

3.1 DATA

The proposed method compares reference data sets to assessment data sets. The reference data set consists of the calibration and validation data used to develop the Chesapeake Bay Index of Biotic Integrity (B-IBI) (Weisberg et al. 1997; Alden et al. 2002). The B-IBI consists of benthic community metrics and scoring thresholds that were developed separately for seven habitat types (Table 1). Table 2 lists the number of original index development samples used in the present study. Weisberg et al. calculated metrics based on averages of three replicate samples per site, while Alden et al. used single replicate samples. We used the metrics values produced by these two studies, but re-calculated B-IBI scores from these metrics to be consistent with the latest B-IBI methodology. The methods for the calculation of the Chesapeake B-IBI are described in the World Wide Web at: <http://www.baybenthos.versar.com/referenc.htm>.

The assessment data set consisted of random samples collected from 1998 to 2002 throughout the Chesapeake Bay. A total of 1,525 samples (single replicates) were used, including 750 samples collected by the Maryland Chesapeake Bay benthic monitoring program, 500 samples collected by the Virginia Chesapeake Bay benthic monitoring program, and 275 samples collected by the Elizabeth River benthic biological monitoring program. Additionally, samples collected from 1996 to 2000 were assembled into a secondary data set, which was evaluated to simulate 305(b) assessments conducted every two years on a running 5-year block of data. All assessment samples were collected with a Young grab (440 cm² surface area, 0.5-mm screen). For sample collection information and methods, see the benthic monitoring program comprehensive reports posted at the World Wide Web address given above.

Assessments were produced for each of 90 Chesapeake Bay Program segments and sub-segments

containing benthic data. Six out of 78 segments and 29 out of 47 additional sub-segments, did not have benthic data. Segments (TMWA 1999) are Chesapeake Bay regions having similar salinity and hydrographic characteristics. In Virginia, segments were sub-divided into smaller units by Virginia DEQ. Sub-segments were produced for each of the main stems of rivers and bays (e.g., James River mesohaline) and for some of the smaller systems opening into the main stem (e.g., Pagan River). Assessment samples were assigned to segments and sub-segments using GIS software. Existing hydrographic data for each sample were used to assign each sample to one of seven B-IBI habitat classes for calculation of the B-IBI (Table 1). These are the same habitat classes used in the reference data set.

3.2 REFERENCE FREQUENCY DISTRIBUTIONS

The number of possible B-IBI outcomes differs by habitat. A greater number of B-IBI scores are possible for habitats with larger number of metrics (e.g., high mesohaline and polyhaline) than habitats with fewer metrics (e.g., tidal freshwater). Some metrics do not score if the metric cannot be calculated (e.g., ratios for which the denominator is zero). This contributes to the disparity of possible B-IBI outcomes among habitats. For this reason, and to allow comparisons among habitats and between the reference and the segment frequency distributions, the B-IBI scores were classified into ordered response categories (“condition categories”) from 1 to 4 as follows:

Condition Category	B-IBI Scores
1	1.0-2.0
2	2.1-2.9
3	3.0-3.9
4	4.0-5.0

The ranges of the condition categories were consistent with the benthic community condition classification system of the Chesapeake Bay benthic monitoring program, except that a “marginal” category was not considered here. The Chesapeake Bay benthic monitoring program classifies scores as follows: meets goal (B-IBI \geq 3.0), marginal (B-IBI = 2.7-2.9), degraded (B-IBI = 2.1-2.6), and severely degraded (B-IBI \leq 2.0).

Examination of reference frequencies distributions with four condition categories (Figure 1) revealed differences among distributions that were statistically significant with a Kolmogorov-Smirnov 2-sided test at the 0.05 probability level (Table 3a). Polyhaline sands differed significantly from all other habitats. Examination of the reference frequency distribution for this habitat (Figure 1) revealed a majority of scores in category 4. Contaminants do not usually accumulate in sandy sediments, so high B-IBI scores would be expected for this habitat. However, the higher B-IBI scores of polyhaline sands can be explained ecologically without having to invoke sediment quality. Polyhaline sands usually harbor greater number of species than polyhaline muds, thus accounting for greater diversity and higher B-IBI scores. To minimize this kind of bias, categories 3 and 4 were combined and the reference frequency distributions re-tested. The combined condition category is thus equivalent to the “meets goal” category of the Chesapeake Bay benthic monitoring program. Using this combined category, additional significant differences were found between the low (tidal fresh and oligohaline) and the high (mesohaline to polyhaline) salinity habitats (Table 3b).

3.3 SEGMENTS WITH MULTIPLE HABITATS

Using the 1998-2002 assessment data, there were a total of 68 segments with more than one habitat (Figure 2). The number of samples in each habitat class within a segment varied, but many segments had very few samples per habitat, often only one sample in a habitat class. Use of the weighted mean approach would have been problematic because of the low sample sizes in some habitats. To see if the number of segments with more than one habitat could be reduced significantly, we combined homogeneous habitats into four habitat strata according to salinity: tidal freshwater, oligohaline, mesohaline, and polyhaline. Even with the combined habitats, there were 49 segments with more than one habitat (Figure 3). Similar results were obtained with the 1996-2000 assessment data (not shown).

3.4 RECOMMENDATION FOR STATISTICAL METHOD

Segment and reference scores represent two independent ordered multinomial distributions. The test of interest is whether these two populations have the same underlying multinomial distribution. However, in the case of multiple habitats, approaches that use stratified data must be considered. Because the habitats had significantly different reference distributions, it was not appropriate to pool distributions across habitats. This limits the choice of a statistical test. Tests based on exact binomial distributions, such as Fisher's Exact (Agresti 1990; Hollander and Wolfe 1999), are not valid for stratified data, and use of a weighted mean approach is problematic because of the low sample size for some of the habitats (some segments had only 1 sample within a specific habitat). A stratified Wilcoxon rank sum test (Lehmann 1998) was found to be most suitable for this type of problem.

3.5 EVALUATION OF DATA SETS AND TYPE I AND II ERRORS

The Wilcoxon rank sum test was applied using Proc-StatXact 5 (Cytel Software Corporation 2002) running under SAS. B-IBI scores were grouped into three condition categories (1, 2, and 3&4 combined, see above) and the distribution of scores within a segment was compared to the reference distribution, treating the scores in each population of samples as ordered categorical responses. Under the null hypothesis (H_0) of no impairment, the two populations (segment and reference) can be considered to have the same underlying multinomial distributions of samples among the ordered categories. The assessment of impairment was based on a one-sided exact test of H_0 against the alternative hypothesis that the segment had a distribution shifted towards lower B-IBI scores than for the reference condition. For segments with multiple habitats, it was assumed that the samples in both populations were random within each habitat class. In the stratified Wilcoxon rank sum test, the ranking is done separately by habitat, and then combined across habitats. The strata weights are based on the frequency of samples in the habitats. Because samples in the Chesapeake Bay benthic monitoring program are allocated randomly within each segment, the number of samples in each habitat is, on average, proportional to the area of the habitat.

Table 4 shows the results of the test for the 1998-2002 assessment data and seven habitats. For each

segment, p-values and the power of the test are presented and the segment is categorized as “pass” (H_0 accepted) or “fail” (H_0 rejected) for each of 3 alternative alpha (0.01, 0.05, and 0.1) and power (any, 0.8, and 0.9) levels. Table 5 presents the results for the four salinity habitats. Table 6 presents a summary of the results, with the number of segments failing for each alternative alpha and power levels and each habitat stratification (i.e., 7 vs. 4 habitats). As expected, the number of segments failing increased with increasing alpha and decreasing power. The choice of 7 vs. 4 habitats made little difference in the number of segments failing. For a power of 90% or greater and a 1% probability level, 38 segments failed for both habitat stratifications. This was the minimum number of segments failing among all of the alpha and power options considered. The 38 segments failing generally matched a-priori expectations of the data (e.g., segments with strong dissolved oxygen or contaminant problems would be expected to differ significantly from reference conditions). Therefore, it is recommended that a segment be considered impaired if the downward shift in B-IBI scores is significant at the $\alpha = 0.01$ level, and with the additional condition that the test has a power of 0.9 or greater. This latter requirement controls for type II errors. For the 1996-2000 data, there were 32 segments failing using this approach, and only 17% of the segments assessed (14 out of 84 segments) had fail/pass classifications that were different from those resulting from the 1998-2002 data analysis.

3.6 MINIMUM REQUIRED SAMPLE SIZE

Because an exact stratified Wilcoxon test was employed for testing impairment, there is no required minimum sample size for testing a segment. Nevertheless, we recommend that segments with less than 10 samples not be tested for impairment. In general, the recommended alpha level of 1% and 90% power has the effect of producing results that are not significant for segments with low sample sizes. Also, a minimum sample size of 10 is consistent with the biocriteria framework for non-tidal streams, where 10 or more samples are needed to evaluate impairment in 8-digit watersheds. The power for a specified sample size cannot be calculated because of the stratification. Power for a given sample size varies according to how the samples are distributed among the strata, and thus it cannot be calculated in advance.

3.7 FINAL ASSESSMENT

The final assessment, based on the 1998-2002 data set and B-IBI scores grouped into three condition categories and seven habitat strata, used the following requirements: an alpha level of 1%, 90% power, and a minimum sample size for the segment of 10. Based on these requirements, 26 segments were considered impaired (Table 7). Impaired segments were then sorted according to how many samples in the segment had B-IBI scores less than 3.0 (the “meets goal” threshold of the benthic monitoring program) compared to the reference condition (Table 7). The proportion of samples in the segment with B-IBI scores less than 3.0 was calculated as the sum of weighted proportions in each habitat (p_i). The weights (w_i) applied to each habitat were calculated as the number of samples in the habitat divided by the total number of samples in the segment. For the reference and degraded distributions, proportions were calculated as above and the segment weights were applied. Thus, the Southern Branch of the Elizabeth River had the largest proportion of failing samples and the largest difference with the reference distribution (Table 7). Segment CB7PHa of the Virginia mainstem had the lowest proportion of failing samples and the smallest difference. Eighty

percent of all B-IBI scores in segment CB7PHA were equal to or greater than 3.0. The difference in failing scores between the segment and the reference distributions can then be used as a guideline of the relative impairment of the segment. Thus, CB7PHA can be considered as having only minor impairment. Applying a Hodges-Lehmann procedure for pooled (un-stratified) data for this segment revealed only a 0.33 B-IBI unit shift between the segment and the reference condition.

4.0 RECOMMENDATIONS

The stratified Wilcoxon rank sum test used in this study is most suitable for stratified data, that is, data segregated into multiple habitats for which the reference distributions are not homogeneous. One limitation of this approach is that it does not allow estimation of the magnitude of shift (e.g., with a Hodges-Lehmann confidence interval). Thus, the distribution of scores in a segment may be significantly different from the reference distribution, but we cannot estimate the magnitude of this difference when the scores are stratified by habitat. The stratified Wilcoxon rank sum test is sensitive to small shifts in distribution of scores. Thus, caution should be used when interpreting the results. Since the magnitude of the shift cannot be estimated, it is recommended that best professional judgment or additional guidelines (see Section 3.7) be used to determine impairment. The reference sites used in the B-IBI index development are the “best of the best”, and may not be representative of the typical distribution of scores for good benthic community condition.

Further research into alternative methods is recommended. Future assessments should take into account the magnitude of the difference between the segment and the reference distribution, and determine what an ecological meaningful difference should be.

5.0 REFERENCES

- Agresti, A. 1990. *Categorical Data Analysis*. John Wiley & Sons.
- Alden, R.A. III, D.M. Dauer, J.A. Ranasinghe, L.C. Scott, and R.J. Llansó. 2002. Statistical verification of the Chesapeake Bay benthic index of biotic integrity. *Environmetrics* 13:473-498.
- Cochran, W.G. 1977. *Sampling Techniques*. 3rd edition. John Wiley & Sons.
- Cytel Software Corporation. 2002. *Pro-StatXact for SAS users. Statistical Software for Exact Non-Parametric Inference*.
- Hollander, M. and D.A. Wolfe. 1999. *Nonparametric Statistical Methods*. John Wiley & Sons.
- Lehmann, E.L. 1998. *Nonparametrics. Statistical Methods Based on Ranks*. Revised 1st Edition. Prentice Hall.
- TMWA (Tidal Monitoring and Analysis Workgroup). 1999. *Chesapeake Bay Program, Analytical*

Segmentation Scheme for the 1997 Re-evaluation and Beyond. Prepared for the U.S. Environmental Protection Agency, Chesapeake Bay Program Office, by the Tidal Monitoring and Analysis Workgroup of the Chesapeake Bay Program Monitoring and Assessment Subcommittee, Annapolis, MD.

Weisberg, S.B., J.A. Ranasinghe, D.M. Dauer, L.C., Schaffner, R.J. Diaz, and J.B. Frithsen. 1997. An estuarine benthic index of biotic integrity (B-IBI) for Chesapeake Bay. *Estuaries* 20:149-158.

Table 1. Habitat classification for the Chesapeake Bay B-IBI. Shown in parentheses are habitat designation numbers used in all further data presentations.

Habitat Class	Bottom Salinity (psu)	Silt-clay (<62 μ) content by Weight (%)
Tidal freshwater (1)	0-0.5	N/A
Oligohaline (2)	$\geq 0.5-5$	N/A
Low mesohaline (3)	$\geq 5-12$	N/A
High mesohaline sand (41)	$\geq 12-18$	0-40
High mesohaline mud (42)	$\geq 12-18$	>40
Polyhaline sand (51)	≥ 18	0-40
Polyhaline mud (52)	≥ 18	>40

Table 2. Number of samples by habitat in the original index development data file used in Weisberg et al. (1997) and Alden et al. (2002). Calibration (Cal) and validation (Val) samples combined. See Table 1 for habitat designation numbers.

	Habitat						
	1	2	3	41	42	51	52
Cal + Val							
Degraded	136	92	49	5	81	7	136
Reference	75	32	20	14	39	39	24

Table 3. Reference frequency distribution comparisons among habitats for (A) B-IBI values grouped into four condition classes and (B) condition classes 3 & 4 combined. Kolmogorov-Smirnov 2-sided test, X = $p < 0.05$. See Table 1 for habitat designation numbers.

A		B-IBI Cond. Classes: 1,2,3,4							
		Habitat Class							
		1	2	3	41	42	51	52	
Habitat Class	1						X	X	
	2						X		
	3	X					X		
	41						X		
	42						X		
	51	X	X	X	X	X		X	
	52	X					X		

B		B-IBI Cond. Classes: 1,2,3&4 Combined							
		Habitat Class							
		1	2	3	41	42	51	52	
Habitat Class	1			X	X	X	X	X	
	2						X		
	3	X							
	41	X							
	42	X							
	51	X	X						
	52	X							

Table 4. Wilcoxon rank sum test results for 90 Chesapeake Bay segments and sub-segments using 1998-2002 assessment B-IBI scores and seven habitats. Shown are power values, p values, and sample sizes by habitat. Segments are categorized as “pass” or “fail” for each of three different alpha (0.01, 0.05, and 0.1) and power (any, 0.8, and 0.9) levels. 0 = “pass”, 1 = “fail”. Habitat codes as in Table 1.

Segment	betaH1=.1	p-value	p-level, any power			p-level, and power ≥.8			p-level, and power ≥.9			Sample sizes by habitat							
	power		0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	1	2	3	41	42	51	52	all
BACOH	0.91	0.0001	1	1	1	1	1	1	1	1	1		3	2					5
BIGMH	1.00	0.0005	1	1	1	1	1	1	1	1	1				1	4	2	1	8
BOHOH	0.07	0.0130	0	1	1	0	0	0	0	0	0			2				2	
BSHOH	0.99	0.0358	0	1	1	0	1	1	0	1	1		6	3				9	
CB1TF	1.00	0.0241	0	1	1	0	1	1	0	1	1	8	8	2				18	
CB2OH	1.00	0.4095	0	0	0	0	0	0	0	0	0		6	35	1	6		48	
CB3MH	1.00	0.0000	1	1	1	1	1	1	1	1	1			20	10	24		55	
CB4MH	1.00	0.0000	1	1	1	1	1	1	1	1	1			2	15	10	1	30	
CB5MH	1.00	0.0000	1	1	1	1	1	1	1	1	1				16	9	17	46	
CB6PHa	1.00	0.0199	0	1	1	0	1	1	0	1	1				1		9	17	
CB7PHa	1.00	0.0068	1	1	1	1	1	1	1	1	1				1		34	41	
CB8PHa	0.58	0.1834	0	0	0	0	0	0	0	0	0					15		15	
CHKOHa	0.97	0.5229	0	0	0	0	0	0	0	0	0	1	1	1				3	
CHOMH1	0.99	0.0007	1	1	1	1	1	1	1	1	1				4	5		9	
CHOMH2	1.00	0.0026	1	1	1	1	1	1	1	1	1			9	3	2		14	
CHOOH	0.54	0.0648	0	0	1	0	0	0	0	0	0		2	3				5	
CHOTF	0.00	0.2121	0	0	0	0	0	0	0	0	0		1					1	
CHSMH	1.00	0.0003	1	1	1	1	1	1	1	1	1			14	12	9		35	
CHSOH	0.54	0.3856	0	0	0	0	0	0	0	0	0		1	3				4	
CHSTF	0.12	0.0476	0	1	1	0	0	0	0	0	0			1				1	
CRRMHa	1.00	0.0061	1	1	1	1	1	1	1	1	1				1	5	1	7	
EASMH	0.88	0.0061	1	1	1	1	1	1	0	0	0				4	2		6	
EBEMHa	1.00	0.0000	1	1	1	1	1	1	1	1	1					17	6	32	
ELIMHa	1.00	0.0000	1	1	1	1	1	1	1	1	1				1		19	36	
ELIPHa	1.00	0.0009	1	1	1	1	1	1	1	1	1					9	16	25	
ELKOH	0.93	0.3435	0	0	0	0	0	0	0	0	0		6	1				7	
FSBMH	0.24	0.0512	0	0	1	0	0	0	0	0	0			4	1			5	
GUNOH	0.98	0.0074	1	1	1	1	1	1	1	1	1		3	7				10	
HNGMH	1.00	0.0000	1	1	1	1	1	1	1	1	1				3	3		7	
JMSMHa	1.00	0.0000	1	1	1	1	1	1	1	1	1			14	3	5	8	40	

Segment	betaH1=.1	p-value	p-level, any power			p-level, and power $\geq .8$			p-level, and power $\geq .9$			Sample sizes by habitat							
	power		0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	1	2	3	41	42	51	52	all
JMSMHb	1.00	0.0000	1	1	1	1	1	1	1	1	1			1	1	2	2	2	8
JMSMHc	0.53	0.0488	0	1	1	0	0	0	0	0	0				3			3	
JMSMHd	0.62	0.0684	0	0	1	0	0	0	0	0	0				3	1		4	
JMSOHa	1.00	0.0001	1	1	1	1	1	1	1	1	1	2	11	11	2	3		29	
JMSOHc	0.00	0.0952	0	0	1	0	0	0	0	0	0			1				1	
JMSPHa	0.88	0.4236	0	0	0	0	0	0	0	0	0						7	5	12
JMSPHd	0.07	0.0031	1	1	1	0	0	0	0	0	0							2	2
JMSTFa	1.00	0.0685	0	0	1	0	0	1	0	0	1	11	4						15
LAFMHa	1.00	0.0000	1	1	1	1	1	1	1	1	1					9	8	18	35
LCHMH	1.00	0.0000	1	1	1	1	1	1	1	1	1				5	3	1		9
LYNPHa	0.00	0.0800	0	0	1	0	0	0	0	0	0							1	1
MAGMH	1.00	0.0000	1	1	1	1	1	1	1	1	1			9	4	7			20
MANMH	1.00	0.0029	1	1	1	1	1	1	1	1	1				3	3	1	1	8
MATTF	0.00	0.1212	0	0	0	0	0	0	0	0	0		1						1
MIDOH	0.94	0.1859	0	0	0	0	0	0	0	0	0		4	2					6
MOBPHa	0.87	0.0074	1	1	1	1	1	1	0	0	0						6	3	9
MOBPHe	0.28	0.0500	0	0	1	0	0	0	0	0	0						1		1
MOBPHf	0.28	0.0500	0	0	1	0	0	0	0	0	0						1		1
MOBPHg	0.12	0.0400	0	1	1	0	0	0	0	0	0							1	1
MOBPHh	0.40	0.0030	1	1	1	0	0	0	0	0	0						1	1	2
MPNOHa	1.00	0.0276	0	1	1	0	1	1	0	1	1	2	1	3	1	1	2		10
MPNTFa	0.94	0.2105	0	0	0	0	0	0	0	0	0	4							4
NANMH	1.00	0.0051	1	1	1	1	1	1	1	1	1		1	5	1	4			11
NANOH	0.91	0.0424	0	1	1	0	1	1	0	1	1		3	3					6
NORTF	1.00	0.0163	0	1	1	0	1	1	0	1	1	1	4						5
PATMH	1.00	0.0000	1	1	1	1	1	1	1	1	1			26	5	12		2	45
PAXMH	1.00	0.0000	1	1	1	1	1	1	1	1	1			18	26	64			108
PAXOH	1.00	0.0209	0	1	1	0	1	1	0	1	1	1	6	5		2			14
PAXTF	0.98	0.0063	1	1	1	1	1	1	1	1	1	2	1						3
PIAMHa	0.12	0.0400	0	1	1	0	0	0	0	0	0							1	1
PMKOHa	1.00	0.0025	1	1	1	1	1	1	1	1	1	1	4	2				2	9
PMKTFa	0.99	0.5801	0	0	0	0	0	0	0	0	0	1	2						3
POCMH	1.00	0.0000	1	1	1	1	1	1	1	1	1					5	5	1	11

Segment	betaH1=.1	p-value	p-level, any power			p-level, and power $\geq .8$			p-level, and power $\geq .9$			Sample sizes by habitat							
	power		0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	1	2	3	41	42	51	52	all
POCOH	0.96	0.0377	0	1	1	0	1	1	0	1	1			4		2			6
POTMH	1.00	0.0000	1	1	1	1	1	1	1	1	1			23	23	45	2	5	98
POTOH	1.00	0.0007	1	1	1	1	1	1	1	1	1	1	6	8				15	
POTTF	1.00	0.0142	0	1	1	0	1	1	0	1	1	6	5					11	
RHDMH	0.73	0.1840	0	0	0	0	0	0	0	0	0			2		1		3	
RPPMHa	1.00	0.0000	1	1	1	1	1	1	1	1	1		2	17	27	26	12	12	96
RPPMHc	0.00	0.0667	0	0	1	0	0	0	0	0	0				1			1	
RPPMHd	0.12	0.0400	0	1	1	0	0	0	0	0	0						1	1	
RPPMHf	0.00	0.1250	0	0	0	0	0	0	0	0	0					1		1	
RPPMHh	0.00	0.1250	0	0	0	0	0	0	0	0	0					1		1	
RPPMHm	0.00	0.6711	0	0	0	0	0	0	0	0	0	1						1	
RPPOHa	1.00	0.0096	1	1	1	1	1	1	1	1	1	1	5	1				7	
RPPTFa	1.00	0.3555	0	0	0	0	0	0	0	0	0	6	4					10	
SASOH	0.94	0.0198	0	1	1	0	1	1	0	1	1		5	2				7	
SBEMHa	1.00	0.0000	1	1	1	1	1	1	1	1	1			2	7	6	38	63	116
SEVMH	1.00	0.0196	0	1	1	0	1	1	0	1	1			5		8		13	
SOUMH	1.00	0.0009	1	1	1	1	1	1	1	1	1			3	3	2		8	
TANMH	1.00	0.0000	1	1	1	1	1	1	1	1	1				19	2	11	6	38
WBEMHa	1.00	0.0000	1	1	1	1	1	1	1	1	1					1	9	29	39
WICMH	1.00	0.2696	0	0	0	0	0	0	0	0	0		1	5	3	1		10	
WSTMH	0.85	0.1169	0	0	0	0	0	0	0	0	0			3		2		5	
YRKMHa	1.00	0.0000	1	1	1	1	1	1	1	1	1			1	9	21	12	23	66
YRKMHb	0.00	0.1250	0	0	0	0	0	0	0	0	0					1		1	
YRKMHc	0.35	0.1267	0	0	0	0	0	0	0	0	0				1	1		2	
YRKPHa	1.00	0.0000	1	1	1	1	1	1	1	1	1						12	15	27
YRKPHd	0.07	0.0031	1	1	1	0	0	0	0	0	0							2	2
YRKPHe	0.00	0.0800	0	0	1	0	0	0	0	0	0							1	1

Table 5. Wilcoxon rank sum test results for 90 Chesapeake Bay segments and sub-segments using 1998-2002 assessment B-IBI scores and four habitat classes. Segments categorized as in Table 4. Habitat codes: 1 = tidal freshwater, 2 = oligohaline, 3 = mesohaline, 4 = polyhaline.

Segment	betaH1=.1	p-value	p-level, any power			p-level, and power>=.8			p-level, and power>=.9			Sample sizes by habitat				
	power		0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	1	2	3	4	all
BACOH	1.00	0.0001	1	1	1	1	1	1	1	1	1		3	2		5
BIGMH	1.00	0.0002	1	1	1	1	1	1	1	1	1			5	3	8
BOHOH	0.71	0.0130	0	1	1	0	0	0	0	0	0			2		2
BSHOH	1.00	0.0454	0	1	1	0	1	1	0	1	1		6	3		9
CB1TF	1.00	0.0276	0	1	1	0	1	1	0	1	1	8	8	2		18
CB2OH	1.00	0.3516	0	0	0	0	0	0	0	0	0		6	42		48
CB3MH	1.00	0.0000	1	1	1	1	1	1	1	1	1			54	1	55
CB4MH	1.00	0.0000	1	1	1	1	1	1	1	1	1			27	3	30
CB5MH	1.00	0.0000	1	1	1	1	1	1	1	1	1			25	21	46
CB6PHa	1.00	0.0200	0	1	1	0	1	1	0	1	1			1	16	17
CB7PHa	1.00	0.0099	1	1	1	1	1	1	1	1	1			1	40	41
CB8PHa	0.96	0.0929	0	0	1	0	0	1	0	0	1				15	15
CHKOHa	1.00	0.4971	0	0	0	0	0	0	0	0	0	1	1	1		3
CHOMH1	1.00	0.0003	1	1	1	1	1	1	1	1	1			9		9
CHOMH2	1.00	0.0053	1	1	1	1	1	1	1	1	1			14		14
CHOOH	1.00	0.1283	0	0	0	0	0	0	0	0	0		2	3		5
CHOTF	0.00	0.2121	0	0	0	0	0	0	0	0	0		1			1
CHSMH	1.00	0.0001	1	1	1	1	1	1	1	1	1			35		35
CHSOH	1.00	0.4167	0	0	0	0	0	0	0	0	0		1	3		4
CHSTF	0.00	0.0676	0	0	1	0	0	0	0	0	0			1		1
CRRMHa	1.00	0.0020	1	1	1	1	1	1	1	1	1			6	1	7
EASMH	1.00	0.0023	1	1	1	1	1	1	1	1	1			6		6
EBEMHa	1.00	0.0000	1	1	1	1	1	1	1	1	1			17	15	32
ELIMHa	1.00	0.0000	1	1	1	1	1	1	1	1	1			1	35	36
ELIPHa	1.00	0.0001	1	1	1	1	1	1	1	1	1				25	25
ELKOH	1.00	0.3490	0	0	0	0	0	0	0	0	0		6	1		7
FSBMH	0.97	0.0842	0	0	1	0	0	1	0	0	1			5		5
GUNOH	1.00	0.0078	1	1	1	1	1	1	1	1	1		3	7		10
HNGMH	1.00	0.0029	1	1	1	1	1	1	1	1	1			6	1	7

Segment	betaH1=.1	p-value	p-level, any power			p-level, and power>=.8			p-level, and power>=.9			Sample sizes by habitat				
	power		0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	1	2	3	4	all
JMSMHa	1.00	0.0000	1	1	1	1	1	1	1	1	1			22	18	40
JMSMHb	1.00	0.0000	1	1	1	1	1	1	1	1	1			4	4	8
JMSMHc	0.93	0.0155	0	1	1	0	1	1	0	1	1			3		3
JMSMHd	0.96	0.0649	0	0	1	0	0	1	0	0	1			4		4
JMSOHa	1.00	0.0000	1	1	1	1	1	1	1	1	1	2	11	16		29
JMSOHa	0.00	0.1081	0	0	0	0	0	0	0	0	0			1		1
JMSPHa	0.87	0.4119	0	0	0	0	0	0	0	0	0				12	12
JMSPHd	0.33	0.0014	1	1	1	0	0	0	0	0	0				2	2
JMSTFa	1.00	0.0685	0	0	1	0	0	1	0	0	1	11	4			15
LAFMHa	1.00	0.0000	1	1	1	1	1	1	1	1	1			9	26	35
LCHMH	1.00	0.0000	1	1	1	1	1	1	1	1	1			8	1	9
LYNPHa	0.55	0.0469	0	1	1	0	0	0	0	0	0				1	1
MAGMH	1.00	0.0000	1	1	1	1	1	1	1	1	1			20		20
MANMH	1.00	0.0039	1	1	1	1	1	1	1	1	1			6	2	8
MATTF	0.00	0.1212	0	0	0	0	0	0	0	0	0		1			1
MIDOH	1.00	0.2035	0	0	0	0	0	0	0	0	0		4	2		6
MOBPHa	0.94	0.0111	0	1	1	0	1	1	0	1	1				9	9
MOBPHb	0.55	0.0469	0	1	1	0	0	0	0	0	0				1	1
MOBPHc	0.55	0.0469	0	1	1	0	0	0	0	0	0				1	1
MOBPHd	0.55	0.0312	0	1	1	0	0	0	0	0	0				1	1
MOBPHe	0.33	0.0024	1	1	1	0	0	0	0	0	0				2	2
MPNOHa	1.00	0.0349	0	1	1	0	1	1	0	1	1	2	1	5	2	10
MPNTFa	0.94	0.2105	0	0	0	0	0	0	0	0	0	4				4
NANMH	1.00	0.0096	1	1	1	1	1	1	1	1	1		1	10		11
NANOH	1.00	0.0540	0	0	1	0	0	1	0	0	1		3	3		6
NORTF	1.00	0.0163	0	1	1	0	1	1	0	1	1	1	4			5
PATMH	1.00	0.0000	1	1	1	1	1	1	1	1	1			43	2	45
PAXMH	1.00	0.0000	1	1	1	1	1	1	1	1	1			108		108
PAXOH	1.00	0.0216	0	1	1	0	1	1	0	1	1	1	6	7		14
PAXTF	0.98	0.0063	1	1	1	1	1	1	1	1	1	2	1			3
PIAMHa	0.55	0.0312	0	1	1	0	0	0	0	0	0				1	1
PMKOHa	1.00	0.0040	1	1	1	1	1	1	1	1	1	1	4	2	2	9

Segment	betaH1=.1	p-value	p-level, any power			p-level, and power>=.8			p-level, and power>=.9			Sample sizes by habitat				
	power		0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	1	2	3	4	all
PMKTFa	0.99	0.5801	0	0	0	0	0	0	0	0	0	1	2			3
POCMH	1.00	0.0000	1	1	1	1	1	1	1	1	1			5	6	11
POCOH	0.99	0.0245	0	1	1	0	1	1	0	1	1			6		6
POTMH	1.00	0.0000	1	1	1	1	1	1	1	1	1			91	7	98
POTOH	1.00	0.0007	1	1	1	1	1	1	1	1	1	1	6	8		15
POTTF	1.00	0.0142	0	1	1	0	1	1	0	1	1	6	5			11
RHDMH	0.91	0.1899	0	0	0	0	0	0	0	0	0			3		3
RPPMHa	1.00	0.0000	1	1	1	1	1	1	1	1	1		2	70	24	96
RPPMHc	0.00	0.0676	0	0	1	0	0	0	0	0	0			1		1
RPPMHd	0.55	0.0312	0	1	1	0	0	0	0	0	0				1	1
RPPMHf	0.00	0.0676	0	0	1	0	0	0	0	0	0			1		1
RPPMHh	0.00	0.0676	0	0	1	0	0	0	0	0	0			1		1
RPPMHm	0.00	0.6711	0	0	0	0	0	0	0	0	0	1				1
RPPOHa	1.00	0.0122	0	1	1	0	1	1	0	1	1	1	5	1		7
RPPTFa	1.00	0.3555	0	0	0	0	0	0	0	0	0	6	4			10
SASOH	1.00	0.0190	0	1	1	0	1	1	0	1	1		5	2		7
SBEMHa	1.00	0.0000	1	1	1	1	1	1	1	1	1			15	101	116
SEVMH	1.00	0.0115	0	1	1	0	1	1	0	1	1			13		13
SOUMH	1.00	0.0010	1	1	1	1	1	1	1	1	1			8		8
TANMH	1.00	0.0000	1	1	1	1	1	1	1	1	1			21	17	38
WBEMHa	1.00	0.0000	1	1	1	1	1	1	1	1	1			1	38	39
WICMH	1.00	0.4873	0	0	0	0	0	0	0	0	0		1	9		10
WSTMH	0.97	0.0991	0	0	1	0	0	1	0	0	1			5		5
YRKMHa	1.00	0.0000	1	1	1	1	1	1	1	1	1			31	35	66
YRKMHb	0.00	0.0676	0	0	1	0	0	0	0	0	0			1		1
YRKMHc	0.67	0.1308	0	0	0	0	0	0	0	0	0			2		2
YRKPHa	1.00	0.0000	1	1	1	1	1	1	1	1	1				27	27
YRKPHd	0.33	0.0014	1	1	1	0	0	0	0	0	0				2	2
YRKPHe	0.55	0.0469	0	1	1	0	0	0	0	0	0				1	1

Table 6. Summary of Wilcoxon rank sum test results for 90 Chesapeake Bay segments and sub-segments using 1998-2002 assessment B-IBI scores. Shown are number of segments failing for each of three alpha and power levels, for (A) seven habitat classes and (B) four habitat classes.

A

Power		Alpha		
		0.01	0.05	0.1
	0.9	38	49	50
	0.8	40	51	52
	Any	43	60	70

B

Power		Alpha		
		0.01	0.05	0.1
	0.9	38	51	57
	0.8	38	51	57
	Any	41	62	73

Table 7. List of Chesapeake Bay impaired segments showing the habitat-weighted proportion of samples in the segment (Seg) with B-IBI scores less than 3.0 (see Section 3.7), the corresponding weighted proportions for the reference (Ref) and the degraded (Deg) original index development samples, and the difference (Seg-Ref) between the segment and the reference proportions. Segments are sorted according to Seg-Ref, from high to low, with the largest difference between segment and reference B-IBIs listed at the top of the table.

Segment	Name	Sample size	Weighted P less than 3.0			Seg-Ref
			Seg	Ref	Deg	
SBEMHa	Southern Branch Elizabeth River	116	0.93	0.04	0.99	0.89
EBEMHa	Eastern Branch Elizabeth River	32	0.88	0.08	0.98	0.79
WBEMHa	Western Branch Elizabeth River	39	0.82	0.04	0.99	0.78
POTMH	Potomac mesohaline	98	0.81	0.09	0.94	0.72
LAFMHa	Lafayette River	35	0.77	0.06	0.99	0.71
CB4MH	Maryland mainstem	30	0.73	0.09	0.98	0.65
PATMH	Patapsco River	45	0.69	0.07	0.89	0.62
YRKMHa	York River mesohaline	66	0.64	0.07	0.98	0.57
POCMH	Pocomoke River	11	0.64	0.07	0.99	0.56
RPPMHa	Rappahannock River mesohaline	96	0.60	0.08	0.95	0.53
ELIMHa	Elizabeth River mesohaline	36	0.56	0.03	0.99	0.52
CB5MH	Maryland mainstem	46	0.57	0.06	0.99	0.50
JMSMHa	James River mesohaline	40	0.55	0.05	0.93	0.50
YRKPHa	York River polyhaline	27	0.52	0.03	0.99	0.48
POTOH	Potomac River oligohaline	15	0.60	0.12	0.72	0.48
PAXMH	Patuxent River mesohaline	108	0.57	0.10	0.95	0.47
MAGMH	Magothy River	20	0.55	0.08	0.91	0.47
JMSOHa	James River oligohaline	29	0.55	0.13	0.75	0.42
GUNOH	Gunpowder River	10	0.50	0.09	0.75	0.41
TANMH	Tangier Sound	38	0.45	0.06	1.00	0.39
CB3MH	Maryland mainstem	55	0.48	0.10	0.89	0.38
CHOMH2	Choptank River	14	0.43	0.07	0.88	0.36
NANMH	Nanticoke River	11	0.45	0.09	0.87	0.36
CHSMH	Chester River	35	0.43	0.08	0.92	0.35
ELIPHa	Elizabeth River polyhaline	25	0.36	0.04	0.99	0.32
CB7PHa	Virginia mainstem	41	0.20	0.03	1.00	0.17

Figure 1. Frequency distributions of reference B-IBI values by habitat type. B-IBI values grouped into four condition classes: 1 (B-IBI = 1.0-2.0), 2 (B-IBI = 2.1-3.0), 3 (B-IBI = 3.1-3.9), and 4 (B-IBI = 4.0-5.0). See Table 1 for habitat designation numbers.

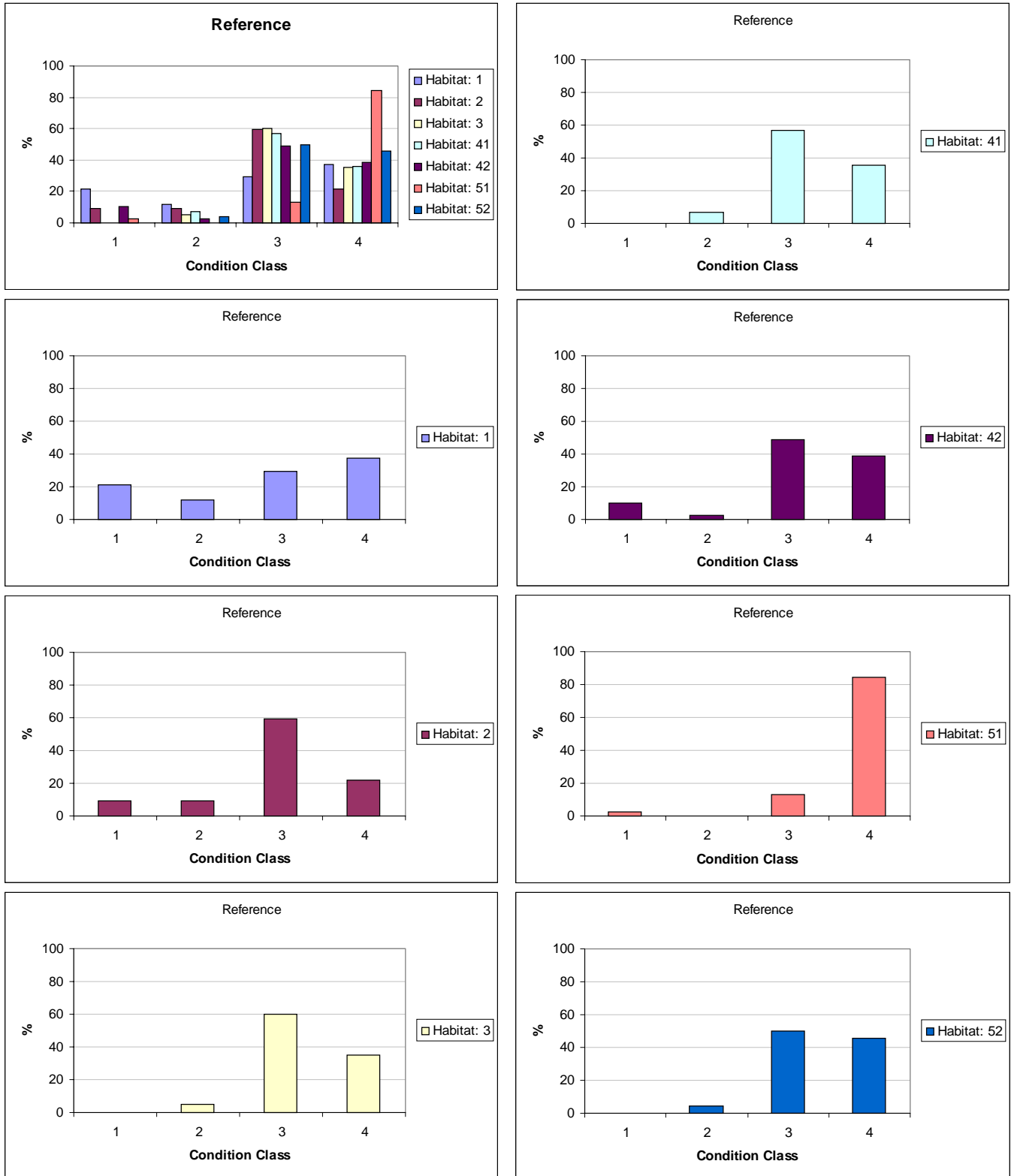


Figure 2. Number of Bay segments with samples in multiple habitats. 1998-2002 assessment, using seven habitat classes (tidal freshwater, oligohaline, low mesohaline, high mesohaline mud, high mesohaline sand, polyhaline mud, and polyhaline sand).

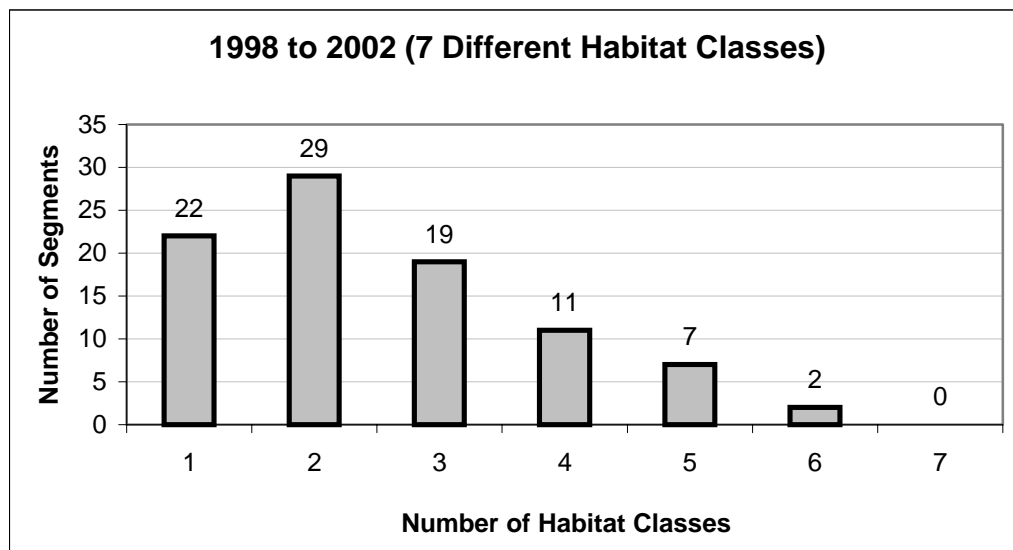


Figure 3. Number of Bay segments with samples in multiple habitats. 1998-2002 assessment, using four habitat classes (tidal freshwater, oligohaline, mesohaline, and polyhaline).

