United States
Environmental Protection
Agency

Proof of Concept for Integrating Bioassessment Results from Three State Probabilistic Monitoring Programs

Final Report

EPA/903/R-05/003
June 2006

Proof of Concept for Integrating Bioassessment Results from Three State Probabilistic Monitoring Programs

Final Report

Prepared for:

Wayne Davis
U.S. Environmental Protection Agency
Office of Environmental Information
Mid-Atlantic Integrated Assessment
Fort Meade, Maryland

COMMITS Contract No. 50CMAA900065

Prepared by:

Versar, Inc.
9200 Rumsey Road
Columbia, Maryland 21045-1934

Recycled /Recyclable
Printed with Vegetable Oil Based Inks on 100%
Post-consumer Process Chlorine Free Recycled Paper

NOTICE

This document has been reviewed and approved in accordance with U.S. Environmental Protection Agency policy. Mention of trade names, products, or services does not convey and should not be interpreted as conveying official EPA approval, endorsement, or recommendation for use.

Funding was provided by the U.S. Environmental Protection Agency under U.S. Department of Commerce, Commerce Information Technical Solutions Contract No. 50-CMAA-900065.

The appropriate citation for this report is:

Southerland, M., Vølstad, J., Erb, L., Weber, E., Rogers, G. 2006. "Proof of Concept for Integrating Bioassessment Results from Three State Probabilistic Monitoring Programs". Report prepared for EPA under Contract No. 50-CMAA-900065. *EPA/903/R-05/003*. U.S. Environmental Protection Agency, Office of Environmental Information and Mid-Atlantic Integrated Assessment Program, Region 3, Ft. Meade, MD.

ACKNOWLEDGEMENTS

Jason Hill (Virginia Department of Environmental Quality), Jeffrey Bailey (West Virginia Department of Environmental Protection), and Maryland Department of Natural Resources provided the data and program information needed to complete this study. Tony Olson (U.S. EPA) provided estimates of stream condition for Virginia and West Virginia. Wayne Davis, Jason Hill, Jeffrey Bailey, Maggie Passmore, Laura Gabanski, and Leska Fore provided valuable comments on the draft report. We greatly appreciate the efforts of Kate Kritcher Traut and Juanita Soto-Smith (Perot Systems Government Services, Inc.) for the technical editing and layout of this report.

ABSTRACT

If data from state stream monitoring assessment programs can be integrated, EPA will be able to obtain estimates of stream condition over larger regions. We assessed the feasibility of integrating three probabilistic monitoring programs—Maryland, Virginia, and West Virginia—and calculated a provisional combined estimate of condition for the non-Coastal Plain region of these states using multimetric indices. All three states had probability-based surveys with similar sample frames (ranges of stream types and sizes) and benthic macroinvertebrate collection procedures outside of the Coastal Plain. Virginia and West Virginia used similar Stream Condition Indices (SCIs) where index scores were derived from the range of values at all sample sites (with thresholds for rating stream condition based on reference condition), while Maryland used a Benthic Index of Biotic Integrity (B-IBI) with metric scores assigned relative to reference condition (and thresholds based on the average of metric scores). To compare the three index methods and establish a common benchmark, SCIs were first calculated for Maryland sites using the Virginia and West Virginia methods. The two SCIs produced nearly identical results on Maryland data indicating that the Virginia and West Virginia methods were directly comparable. The Maryland B-IBI had a more uniform distribution of scores than the SCIs and was not directly comparable. The West Virginia procedure for selecting reference sites included site-by-site best professional judgment (BPJ) exclusions that were more restrictive, but which could not be reproduced for other states, so were not included in the provisional integrated assessment. Application of each state's reference criteria to Maryland data (excluding West Virginia's BPJ exclusions) resulted in different suites of reference sites. However, the distributions of reference sites selected were similar, suggesting the different reference sites were of similar stream quality (comparably affected by human disturbance). Using our example integration approach (and treating each state as a stratum) and the 10th percentile of reference sites as a degradation threshold, we estimated that approximately 39% of all streams in the non-Coastal Plain of the three states would be classified as degraded for 1997-2003. Applying a threshold of degradation derived from higher quality reference sites (e.g., those including West Virginia's BPJ exclusions) would increase the proportion of streams designated as degraded. We conclude that similar integrations at the level of stream condition assessment will be possible even when data integration is problematic.

CONTENTS

Section 1: Introduction 1

Section 2: Summary of State Programs 3
- **2.1:** Maryland
- **2.2:** Virginia
- **2.3:** West Virginia

Section 3: Comparison of Sample Frames, Survey Designs, and Data Collection 10

Section 4: Comparison of Indicators and Reference Conditions 11

Section 5: Integration of Assessments from the Three States 20

Section 6: Discussion and Recommendations 23

Section 7: Literature Cited 26

FIGURES

Figure 1: Calculation of SCI scores 8

Figure 2: Venn diagram of reference sites selected from Maryland data using the Virginia, West Virginia, and Maryland reference criteria 16

Figure 3: Distribution of SCI scores for reference sites selected from Maryland data using Virginia, West Virginia, and Maryland methods 17

Figure 4: Cumulative distribution of Maryland benthic IBI scores and Virginia and West Virginia SCI scores for Maryland streams 19

Figure 5: Cumulative distribution of SCI scores for Maryland, Virginia, and West Virginia, and all three states combined 20

Figure 6: Distribution of SCI scores for West Virginia reference sites selected using all criteria 21

Figure 7: Cumulative distribution of SCI scores for all three states combined at two different thresholds 22

TABLES

Table 1: Comparison of sample frames and survey designs used by stream monitoring programs in Maryland, Virginia, and West Virginia 3

Table 2: Comparison of benthic sampling methods used by stream monitoring programs in Maryland, Virginia, and West Virginia 4

Table 3: Metrics used by each state to create benthic macroinvertebrate IBI or SCI scores 5

Table 4: Criteria used by Maryland, Virginia, and West Virginia to select reference sites for multimetric indices of biological condition based on benthic macroinvertebrates 12

Table 5: Total Habitat score calculation used by Virginia and West Virginia for high and low gradient streams, along with the surrogate applied to Maryland..... 14

1. Introduction

The U.S. Environmental Protection Agency (EPA) is evaluating methods for developing a national assessment of stream conditions. One method for completing such an assessment is to conduct a regional survey of wadeable streams using a probability-based design and standardized sampling protocols such as the Environmental Monitoring and Assessment Program (U.S. EPA 2002a). However, such an effort would be costly and partially redundant with the stream assessment programs that individual states now conduct to characterize water quality (U.S. EPA 2002b). If the methods and results from these state programs are similar enough, EPA may be able to use data already collected by states to assess stream conditions over larger, multi-state regions.

Stream assessment programs from individual states must meet several requirements before they can be used for larger scale assessments. First, each program must address comparable sample frames, i.e., each state must sample a stream network with a comparable range of stream sizes and types, over a similar time period. Second, each state program must have probability-based survey design that allows unbiased area-wide estimates of stream condition to be made with quantifiable precision. Third, each state must have reliable field collection techniques, laboratory protocols, and quality assurance and control procedures that ensure accurate data. Ideally, these methods will have documented performance characteristics (U.S. EPA 2000, NWQMC 2001). Lastly, each state program must have data to support a single metric, set of metrics, or model that can be used to evaluate stream condition (i.e., an assessment endpoint or indicator). Multimetric indices of biological assemblages that characterize stream condition as a single value (Karr 1991, Barbour et al. 1995) are easily interpreted and used by almost every state (U.S. EPA 2002b).

If the above requirements are met, it is still likely that most state programs will differ significantly in the field collection protocols and indicators they use. These differences can prevent directly integrating data into a consolidated data set, but they are unlikely to preclude combining results at the level of stream condition assessment. That is, states may share assessment comparability but not data comparability. For example, two states may employ different sampling gear so that different invertebrate taxa are targeted (e.g., different numbers of mayflies would be collected by each method at the same site), while the indicators used by each state would rate the site in the same condition. This is the virtue of using the deviation from reference condition to rate streams (i.e., indicators with different scores at the same site will be similar distances from reference scores developed for each indicator). Also, it is likely that states will have different survey designs; however, if each is probability-based, the differences will not affect integration at the assessment level, as each state is a de facto stratum that can be combined into a single estimate of the proportion of degraded stream miles.

Little previous work has been done to integrate results of state stream assessment programs over larger regions because few programs have conducted probabilistic surveys that meet the requirements for integration described above. However, Maryland, Virginia, and West Virginia each conducted probability-based surveys of wadeable streams that used benthic macroinvertebrate multimetric indices to evaluate stream condition statewide over 4- or 5-year periods from 1997 to 2004. Here we assess the feasibility of integrating results from these programs for the non-Coastal Plain (Piedmont and Highland) regions of these states, and we describe the steps necessary to achieve a regional assessment of stream condition. Specifically, we

- Compared the sample frames and survey designs of the three states;
- Compared the benthic macroinvertebrate sampling methods of the three states;
- Compared the construction and scoring of the multimetric indices of the three states (by applying all three indices to Maryland 2000-2004 data);
- Compared the criteria used by the three states to select reference sites and the distribution of index scores for each state's reference sites (again, applied to Maryland data);
- Combined the site results from all three states using comparable index scores to produce a regional (non-Coastal Plain) cumulative distribution of scores; and,
- Applied example thresholds of degradation to estimate the regional proportion of stream miles rated as 'good,' 'fair,' or 'poor' (based on different assumptions about reference condition).

2. Summary of State Programs

The relevant components of the stream assessment programs of Maryland, Virginia, and West Virginia are described below. Note that this section describes the three state programs when this study was initiated. All three programs continue to evolve and have already incorporated refinements that are not captured here. Therefore, this study should be viewed as a demonstration of integration principles and not a critique of individual state programs. A comparison of components from each state program is provided for (1) sample frames and survey designs (Table 1), (2) benthic macroinvertebrate sampling methods (Table 2), and (3) metrics used in each state's multimetric index (Table 3).

Table 1. Comparison of sample frames and survey designs used by stream monitoring programs in Maryland, Virginia, and West Virginia.

Program Components	STATE		
	Maryland (2000-2004)	**Virginia (2000-2003)**	**West Virginia (1997-2001)**
Sample Frame	U.S. Geological Survey 1:100,000 stream network 1st through 4th order[1] streams	U.S. EPA RF3 reach file 1:100,000 stream network 1st through 6th order streams	U.S. EPA RF3 reach file 1:100,000 stream network 1st through 5th order streams
Sample Unit	75-m reach	30 to 400-m reaches	100-m reach
Survey Design	Probabilistic (Lattice sampling)	Probabilistic (GRTS[2] design)	Probabilistic (GRTS design)
Survey Density	Target minimum of 10 sites per PSU[3] plus 3-11 additional samples for PSUs with more than 100 stream miles Approximately 300 sites per year 1,500 sites statewide	Target 60 sites per stratum over 5 years with a minimum of 50 sites per year 250 sites statewide	150 sites per year 750 sites statewide
Site Revisits	Same 25 sentinel sites sampled each year	One site per region chosen randomly from all sites sampled to date and revisited per year	No revisits

[1] All states use Strahler (1957) stream order classifications

[2] Generalized random tessellation stratified (Stevens 1997, Stevens and Olsen 2004)

[3] Primary sampling units

Table 2. Comparison of benthic sampling methods used by stream monitoring programs in Maryland, Virginia, and West Virginia.

	MD	VA (SCI non-Coastal Plain)	WV (SCI)
Benthic Field Sampling	600-µm, 0.3-m D-frame net 20 jabs of 1 ft^2	600-µm, 0.3-m D-frame net or 2 m^2 kick net[1] Approximately 2 m^2 total	600-µm, 0.3-m D-frame net or 2 m^2 kick net Approximately 2 m^2 total, eight 0.25 m individual kicks
Field QA	Duplicate samples at 12 to 15 sites per year (7% of all sites)	Duplicated 10% of probabilistic sites during 2001-2004[2]	Duplicate samples at 12 to 15 sites per year
Benthic Habitat Sampled	Multi-habitat Primarily riffles but also rootwads/woody debris/leaf packs, macrophytes, and undercut banks	Single or Multi-habitat One to three kicks per riffle, multi-habitat samples when riffles were rare Samples from downstream half of reach	Single or Multi-habitat
Index Period for Benthos	Approximately March 1 to May 1	Approximately March 1 to May 1	Mid-April through October
Laboratory Methods	Random subsample of approximately 100 organisms based on grid cells Identification to genus or lowest practical taxon (chironomids/ oligochaetes to family)	Random subsample of minimum of 100 organisms or 4 quadrats based on grid cells (2-inch square grids in a 50-quadrat box) Identification to family level	Random subsample of approximately 200 organisms based on grid cells Identification to family level
Laboratory QA	Resample and identification of every 20th sample (7%)	No resampling	Resample and identification of 5% of samples
Benthic Indicator	Benthic Index of Biotic Integrity on 1 to 5 scale	Virginia Stream Condition Index on 0-100 point scale	West Virginia Stream Condition Index on 0-100 point scale

[1] Depending on the amount of riffle habitat available (Barbour et al. 1999)

[2] Jason Hill, Virginia Department of Environmental Quality, personal communication

Table 3. Metrics used by each state to create benthic macroinvertebrate IBI or SCI scores. Lines beginning with a "%" symbol indicate the percentage of that taxon in the total sample. Complete descriptions of the Maryland B-BIBIs, Virginia SCI, and West Virginia SCI are reported in Southerland et al. (2005), Burton and Gerritsen (2003), and Tetra Tech, Inc. (2000), respectively.

Metric Type	MD (B-IBIs for Highlands and Eastern Piedmont)[1]	VA (SCI for non-Coastal Plain)	WV (SCI)
Taxonomic Richness	Number of taxa (genera) Number of EPT[2] taxa Number of Ephemeropteran taxa	Number of taxa (families) Number EPT families	Number of taxa (families) Number of EPT families
Taxonomic Composition	% Chironomidae % Clingers % Tanytarsini % Scrapers % Swimmers % Diptera	% Ephemeroptera % Plecoptera + Trichoptera + Hydropsychidae % Chironomidae % Top 2 Dominant Taxa	%EPT % Chironomidae % Top 2 dominant taxa
Tolerance	% Intolerant to urban stressors % Tolerant taxa % Collectors	HBI[3] % Scrapers	HBI

[1] List of 10 metrics includes those found in either the Highlands B-IBI (8 metrics) or Eastern Piedmont B-IBI (6 metrics)

[2] Ephemeropteran, Plecopteran, or Tricopteran

[3] The Hilsenhoff Biotic Index (HBI) was defined as abundance-weighted average tolerance of assemblage of organisms (family taxonomic level)

2.1 Maryland

The Maryland Biological Stream Survey (MBSS) is a long-term program conducted by the Maryland Department of Natural Resources to assess the condition of the state's nontidal, freshwater streams (Klauda et al. 1998). Benthic macroinvertebrates are collected during spring each year as part of a larger sampling effort and used to calculate a Benthic Index of Biotic Integrity (B-IBI) for Maryland streams. The MBSS has completed two rounds of statewide sampling; the first was conducted in 1995-1997 and the second in 2000-2004. In this study, we used data collected at 596 randomly selected, non-Coastal Plain sites sampled in 2000-2004 to assess stream condition. We also used data from 144 reference sites sampled in 1995-2004.

Sampling sites for the second round of the MBSS were selected from a 1:100,000-scale stream network using a lattice sampling design (see Cochran 1977) to select watersheds randomly in time and space. Eighty-four primary sampling units (PSUs) consisting of one or more Maryland 8-digit watersheds were sampled over 5 years (Roth et al. 2005). It is worth noting that Maryland's 138 8-digit watersheds, averaging 75 m^2, are different from the 20 U.S. Geological Survey (USGS) 8-digit cataloging units in Maryland which average 500 m^2 (Roth et al. 2002). In principle, the survey design supports the use of the Sen-Yates-Grundy variance estimator of statewide mean stream condition because the selection probability of any stream segment, and the joint selection probabilities of any pair of segments for the entire round, is known and greater than zero. Seventeen PSUs were sampled per year, and two randomly selected PSUs were sampled twice during the 5 years. Each PSU had a target minimum of 10 sites sampled, with an additional 3 to 11 sample sites allocated for PSUs with more than 100 stream miles. Streams were stratified into two groups within a PSU, 1st- or 2nd-order streams (Strahler 1957) and 3rd- or 4th-order streams, unless a stratum would have contained less than 10% of the stream miles in the PSU. In that case, sites were selected within the PSU using simple random sampling. The samples within each PSU were allocated proportionally to stream lengths in the strata, ensuring equal selection probability for all stream segments in the PSU. While sampling random PSUs twice in the lattice design provides temporal information, we pooled the samples for each PSU and analyzed the Maryland data using standard stratified random estimators (Cochran 1977) to simplify this analysis (see Berger 2004).

Benthic macroinvertebrates were collected within 75-m sample segments during spring each year using 600-µm D-frame nets (Roth et al. 2005). Twenty kick net samples were taken at each site from riffle, rootwad, and leaf-pack habitats in approximate proportion to their abundance in the stream segment. For each segment, a random subsample of 100 organisms was identified to genus or the lowest practical taxon level. These data were used to calculate a B-IBI for Maryland streams.

The B-IBI rates streams on a scale of 1 to 5 where scores of 4-5 represent good condition, 3-3.9 represent fair, 2-2.9 represent poor, and 1-1.9 represent very poor. In this study we used both the Highlands and Eastern Piedmont MBSS IBIs to cover the non-Coastal Plain region of interest (Southerland et al. 2005). These B-IBIs include 8 and 6 metrics, respectively, related to the number or percentage of different invertebrate taxa in a sample (Table 3). Each metric was rated as a 1, 3, or 5 depending on how it compared to the distribution of scores from a set of reference sites; these reference sites were selected from the entire MBSS dataset using criteria for sites minimally affected by human activities and representative of Maryland non-Coastal Plain streams. Each metric was scored as a 1 if its value was less than the 10th percentile of the reference values, as a 3 if it was in the 10th to 49th percentiles, and as a 5 if it was equal to or greater than the 50th percentile. Metrics that were expected to increase with stream degradation were scored conversely. The B-IBIs were calculated as the average of the metric scores. A value of 3 is also used as the threshold of degradation for the B-IBI; a B-IBI of 3 corresponds to the 9th percentile of Maryland reference sites.

2.2 Virginia

The Virginia Department of Environmental Quality (VDEQ) biomonitoring and assessment program samples fixed and randomly chosen monitoring sites to meet state and federal water quality monitoring requirements. We analyzed 180 randomly chosen 1st- to 6th-order, non-Coastal Plain streams sampled in 2001-2003 as part of the random portion of the survey known as ProbMon (VDEQ 2003).

The survey design used to select Virginia streams for sampling was a generalized random tessellation stratified (GRTS) design (Stevens 1997, Stevens and Olsen 2004) chosen to ensure that a spatially balanced selection of streams was achieved. This design selects sites randomly by Strahler stream order but assigns a greater probability to selecting higher order streams. This ensures that higher order streams are adequately represented in the sample despite constituting a lower proportion of the total stream miles in the state. A target of 50 stream sites was selected from throughout the state each year using EPA Reach File 3 (RF3) overlaid onto a 1:100,000-scale topographic map.

Benthic macroinvertebrates were collected from stream reaches following EPA Rapid Bioassessment Protocols (Barbour et al. 1999). Sample reaches had lengths 30 times the stream width to a maximum of 100 m for 3rd or lower-ordered streams and 400 m for 4th or higher-ordered streams. Invertebrates were collected using 600-µm kick nets and sampling approximately 2 m^2 of riffle substrate. If no riffle habitat was available, multi-habitat samples were collected with 600-µm D-frame nets following Barbour et al. (1999). Either 100 organisms or 16 in^2 (103-cm^2)

of sample material spread out onto a 200-in^2 (1290-cm^2) sampling tray were enumerated and identified to family.

Invertebrate data were used to calculate the Virginia SCI as described in Burton and Gerritsen (2003). The index consisted of 8 metrics related to the abundance, species composition, and environmental tolerance of invertebrates collected (Table 3). Each metric was standardized to a 100-point scale where 0 represented the worst condition observed, and 100 the best. A score corresponded to its rank between the 5th and 95th percentiles in the distribution of all data collected (not just the reference sites). Extreme values below the 5th percentile or greater than the 95th percentile were assigned 0 or 100, respectively (Figure 1). This practice eliminates the influence of outliers and reduces the effect that different datasets in the future will have on setting the SCI scores. The SCI was calculated as the average of the 8 metrics. The 10th percentile of reference sites (adjusted downward 5 SCI points for the variance in duplicate samples) is used to designate the SCI value as the threshold of degradation.

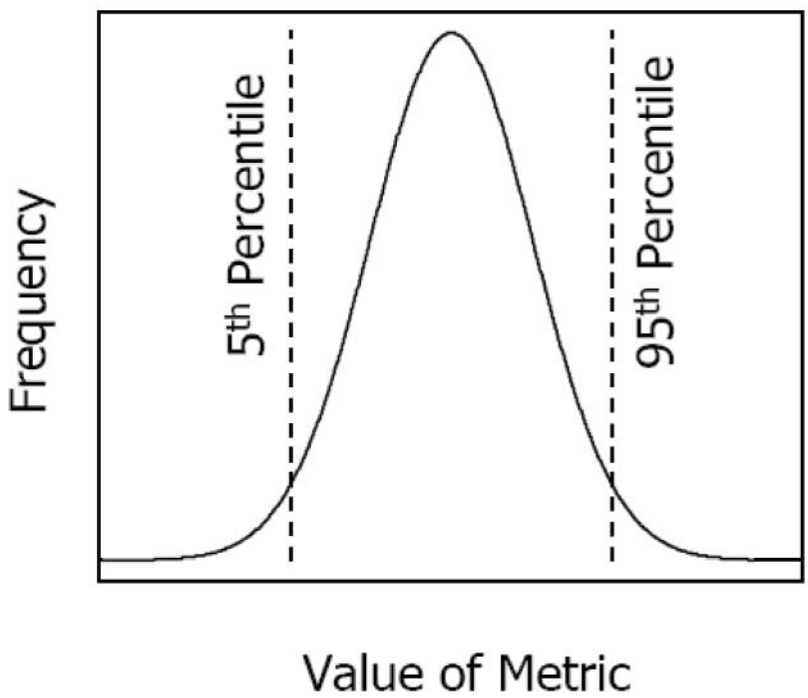

Figure 1. Calculation of SCI scores. All sites (not just reference) are used to determine the range of metric scores. Metric scores are converted to 0-100 scale. For Virginia, the 5th percentile of worst conditions represented 0, and the 95th percentile represented 100. For West Virginia, the procedure was the same except that the worst value represented 0 rather than the 5th percentile. The SCI score was calculated as the mean of the standardized metrics.

2.3 West Virginia

The methods applied by the Watershed Assessment Program of the West Virginia Department of Environmental Protection (WVDEP) for assessing stream condition were similar to those used by Virginia's program. Watersheds were sampled on a rotating basis, completing a statewide survey during 1997-2001. West Virginia used a GRTS survey design to sample streams, but sampled by USGS 8-digit basins. The year of sampling was a stratum in this design and five to seven basins were sampled each year. We used data collected at 716 randomly selected sites from 1st- through 5th-order streams (WVDEP 2005). Invertebrate data were collected according to EPA Rapid Bioassessment Protocols using procedures that differed from those of Virginia by not including multihabitat samples and by having a longer index period (April to October), as described by Tetra Tech, Inc. (2000). Benthic macroinvertebrate data were used to calculate the West Virginia SCI, which is similar to the Virginia SCI except that the index consisted of two fewer metrics (Table 3). Note also that West Virginia scored metrics on a scale of 1-100 based on the 0 to 95th percentiles (Figure 1) rather than the 5th to 95th percentiles used by Virginia. The 5th percentile of reference sites (adjusted downward 7.4 SCI points for the variance in duplicate samples) is used to designate the SCI value as the threshold of degradation.

3. Comparison of Sample Frames, Survey Designs, and Data Collection

The sample frames, survey designs, and data collection used by the three state programs were somewhat different, but comparable (Table 1). Most importantly, all three states used probabilistic survey designs that supported unbiased estimates of means, totals, and proportions statewide with quantifiable precision. Differences in sample frames were small. Maryland sampled no 5th or 6th order streams, but these higher order streams constituted small fractions of the total samples in Virginia (4%) and West Virginia (5%). In addition, the sample frames of all three states were limited to wadeable streams on a 1:100,000-scale stream network, i.e., even the highest order streams sampled were wadeable and therefore not unusually large.

The states used different stream segment lengths as sample units for field data collections. This difference was also expected to have little or no effect on comparability because the stream segment lengths sampled by Maryland (75 m) and West Virginia (100 m) were very similar, as were segments in the lower order streams in Virginia (based on 30 times stream width to a maximum of 100 m on 1st- through 3rd-order streams). The vast majority of streams sampled in Virginia were of lower orders. Overall, the differences in sample frames of the three states were unlikely to cause large differences in assessment results, and did not preclude the calculation of an integrated estimate of stream condition.

Benthic sampling methods were likewise similar among the three states (Table 2). All three states used frame nets to sample, and each focused on riffle habitat where the greatest diversity of invertebrates is expected (Barbour et al. 1999). Laboratory sorting procedures were generally similar, except that (1) Maryland sorted to a lower taxonomic level (genus) than the other states (which sorted to family) and (2) West Virginia sorted a larger sample (200 rather than 100 organisms). In an earlier Maryland study, Vølstad et al. (2003) determined that 200-organism subsamples improved the precision of mean B-IBI scores only marginally over 100-organism subsamples. West Virginia sampled during the summer in addition to spring but did not observe appreciable variation between the two seasons at the family level (Tetra Tech, Inc. 2000). All programs included documented quality assurance procedures.

Based on comparability among sample frames, survey designs, and data collection, the sampling programs of the three states were suitable for conducting an integrated assessment of stream condition.

4. Comparison of Indicators and Reference Conditions

As described earlier, the stream condition indices used by the three states were developed independently. To ultimately integrate the results from the states, we had to (1) obtain a common indicator of stream condition and (2) evaluate its scores in the context of comparable reference conditions. First, we had to determine if the SCIs from Virginia and West Virginia were directly comparable, so that only the Maryland B-IBI (which was most dissimilar) needed to be substituted by an SCI. This would allow us to use the regional estimates already calculated for each state in the final integration. To compare the SCIs, we needed to use a single dataset, so we used Maryland data from 2000-2004 since they were readily available to the investigators. We evaluated the two SCIs by calculating both the Virginia and West Virginia SCI scores for Maryland data, and comparing them to each other. Specifically, the Maryland benthic macroinvertebrate data were reduced from genus to family level identifications for each site because both SCIs were based on family-level identification. The SCI scores were then calculated both using the Virginia SCI (Burton and Gerritsen 2003) and the West Virginia SCI (Tetra Tech, Inc. 2000).

Second, we had to adjust for the different reference conditions used by the three states, which we did by applying each state's reference criteria to Maryland data. Each state used a combination of criteria that measured water chemistry, instream habitat, and land use to select reference sites that represented streams with minimal human disturbance (Tables 4 and 5). All three states used 13 selection criteria, although the specific criteria used varied by state. Maryland was the only state to use percent of forested cover in the watershed or remoteness from human development explicitly, while Virginia and West Virginia used more instream habitat variables than Maryland. Virginia was the only state to use total phosphorus. West Virginia used conductivity and fecal coliform as secondary criteria (i.e., best professional judgment was used to confirm their importance), and applied best professional judgment on a site-by-site basis to identify sites with additional human disturbance.

It is important to remember that both the B-IBI and the SCIs rate stream condition relative to reference conditions, but in different ways. In the Maryland B-IBI, each metric is scored relative to the distribution of reference sites and the average of all metric values is the B-IBI score (see Southerland et al. 2005). Because a metric value of 3 denotes departure from reference, a B-IBI of less than 3 indicates degradation. Thirteen of the 144 Maryland reference sites have scores below 3, which represents the 9th percentile of candidate reference sites. SCI scores are calculated by Virginia and West Virginia based on the distribution of component metric values at all sites sampled (not on reference sites alone as is done for the Maryland B-IBI). The threshold for rating streams as degraded is then applied to the SCI scores as a percentile of SCI scores at reference sites (10th percentile of

Table 4. Criteria used by Maryland, Virginia, and West Virginia to select reference sites for multimetric indices of biological condition based on benthic macroinvertebrates. As part of this study, Stream Condition Indices (SCIs) were calculated for Maryland data using the Virginia and West Virginia methods. Where Maryland data included Virginia and West Virginia reference criteria variables, these criteria were applied directly; in other cases, substitute Maryland variables were used as reference criteria (see two right columns); note that a variable may have substituted for more than one criterion.

		Maryland*	Virginia*	West Virginia*	MD substitute for VA SCI	MD substitute for WV SCI
Chemical Criteria	**pH**	≥ 6	6-9	6-9	6-9	6-9
	DO	≥ 4 ppm	≥ 6 mg/L	≥ 5 mg/L	≥ 6 mg/L	≥ 5 mg/L
	ANC	≥ 50μ eq/L	Not used	Not used	-	-
	Nitrate	≤ 4.2 mg/L	Not used	Not used	-	-
	Conductivity	Not used	< 250 μmhos/cm	< 500 μmhos/cm [secondary criterion]	< 250 μmhos/cm	< 500 μmhos/cm [applied directly even though secondary criterion for WV]
	Fecal Coliform	Not used	Not used	< 800 colonies/100 ml [secondary criterion]	-	< 5% urban [based on strong relationship with urban land use]
	TN	Not used	< 1.5 mg/L	Not used	< 1.5 mg/L	-
	TP	Not used	< 0.05 mg/L	Not used	< 0.05 mg/L	-

(Continued on next page)

* For more information on each state's metrics and criteria, see:

Maryland: Paul et al. 2002; Paul et al. 2003; Roth et al. 2005

Virginia: Burton 2003; VDEQ 2003, 2005

West Virginia: Tetra Tech 2000; West Virginia Department of Environmental Protection 2006

Table 4. Continued

		Maryland	Virginia	West Virginia	MD substitute for VA SCI	MD substitute for WV SCI
Habitat and Land Use Criteria	**% Urban Development**	≤ 5	≤5	Not used	≤	-
	% Forested	≥ 35	Not used	Not used	-	-
	Modified Remoteness Rating	≥ 11	Not used	Not used	-	-
	Aesthetic/Trash Rating	≥ 11	Not used	Not used	-	-
	Instream Habitat Rating	≥ 11	Not used	Not used	-	-
	Anthropogenic Activities/ Disturbances	Not used	BPJ [1]	BPJ [1]	Aesthetics rating ≥11 ≤ 5% urban	Aesthetics rating ≥11 ≤ 5% urban
	Violations of State WQ Standards	Not used	Not used	No violations	-	Nitrate ≤ 4.2 mg/L, ANC ≥ 50µ eq/L, and ≤ 5% urban
	Non-point Pollution	Not used	Not used	None obvious	-	Nitrate ≤4.2 mg/L, ANC ≥ 50µ eq/L, and ≤ 5% urban
	Epifaunal Substrate Score	Not used	≥ 11	≥ 11	Instream habitat score ≥11	Instream habitat score ≥11
	Channel Alteration Score	Not used	≥ 11	≥ 11	Channel alteration score ≥11 and no channelization	Channel alteration score ≥11 and no channelization
	Sediment Deposition Score	Not used	≥ 11	≥ 11	Excluded sites with extensive bar formation	Excluded sites with extensive bar formation
	Bank Disruptive Pressure Score	Not used	≥ 11	≥ 6	Converted bank stability score (≥ 11)[2]	Converted bank stability score (≥ 11)
	Riparian Vegetated Buffer Width Score/m	≥ 30m	≥ 11	≥ 6	≥11 (MD meters converted to 0-20 score)	≥ 6 (MD meters converted to 0-20 score)
	Total Habitat Score	Not used	≥ 140	≥ 130	≥ 108 (scored as described in Table 5)	≥ 108 (scored as described in Table 5)
	Point Source Discharge	No effluent discharge	Not used	Not used [except as part of BPJ]	-	-
	Other	No channelization no storm drains	Not used	Not used [except as part of BPJ]	-	-

[1] Best professional judgment (BPJ) decisions by West Virginia were made in the field based on visual assessment informed by secondary criteria. Virginia used a less formal BPJ to eliminate sites with anthropogenic disturbance. Urban land use ≤ 5% and Maryland aesthetics rating only partially capture this BPJ. See text for additional discussion.

[2] Paul et al. 2002

Table 5. Total Habitat score calculation used by Virginia and West Virginia for high and low gradient streams, along with the surrogate applied to Maryland. Substitutions for variables not measured in Maryland are described in the far right column.

	MD High and Low Gradient Streams	VA/WV High Gradient Streams	VA/WV Low Gradient Streams	MD substitution
Epifaunal Substrate Score	0-20 score	0-20 score	0-20 score	Same
Sediment Deposition Score	Not used	0-20 score	0-20 score	Excluded MBSS sites with extensive bar formation
Channel Flow	Not used	0-20 score	0-20 score	-
Channel Alteration Source	0-20 score	0-20 score	0-20 score	Used a Tetra Tech conversion (0-20)
Bank Disruptive Pressure Score	0-20 score	0-20 score	0-20 score	Used Tetra Tech values to create scores based on SCI method
Riparian Veg (buffer) Zone Width Score	0-20 score	0-20 score	0-20 score	Created a score based on SCI scoring method
Vegetation Protection	Not used	0-20 score	0-20 score	-
Embeddedness	0-20 score	0-20 score	Not used	Created a score based on SCI scoring method
Velocity/Depth	0-20 score	0-20 score	Not used	-
Frequency of Riffles	0-20 score	0-20 score	Not used	Used MBSS riffle quality score (0-20)
Pool Substrate Characterization	0-20 score	Not used	0-20 score	Used MBSS pool quality score (0-20) in place of both pool substrate and pool variability
Pool Variability		Not used	0-20 score	
Channel Sinuosity	0-20 score	Not used	0-20 score	Created a score based on SCI scoring method
Total Possible Points	***180***	***200***	***200***	-

Virginia and 5th percentile for West Virginia, both adjusted downward to account for variability in duplicate samples) that also denotes departure from reference condition. In this way, the threshold of degradation for SCIs is applied to a specific point on the distribution of reference sites. For this reason, differences in reference criteria did not affect the SCI scores calculated by Virginia and West Virginia, allowing us to address comparability of reference condition after the SCI scores were calculated.

The Maryland data included sample values for all chemical criteria used by Virginia and West Virginia, but did not include all habitat variables used by these states as reference criteria. Several variables related to habitat were approximated using similar characteristics collected by the MBSS. We could not construct a surrogate for the BPJ decisions made by West Virginia (by definition), so they were not included in the reference criteria for these analyses (resulting in more reference sites being selected). We will return to the issue of including BPJ in reference criteria later in this report. As described above, rating stream condition with the SCIs is based on applying threshold scores that are percentiles of reference sites, e.g., the boundary between "not degraded" and "degraded" stream conditions. A standard percentile may be used to set this boundary (e.g., 10th percentile for Virginia and 5th percentile for West Virginia) or a set index score (based on average of reference-based metrics) that corresponds to a percentile of reference may be used (e.g., 9th percentile for Maryland that corresponds to a B-IBI of 3). All three states also use a confidence interval around the threshold value for degradation at individual sites to make impairment decisions. These confidence intervals are based on the variability in values from duplicate samples within sites (i.e., the threshold for designating degradation is 5 to 8% lower than the score corresponding to the percentiles listed above). For purposes of illustration, we have used the exact percentile (e.g., 10th) as the threshold for designating streams as not degraded (those with scores above the percentile) and degraded (those with scores below the percentile), without accounting for the uncertainty associated with sample variability within or between sites.

We applied each state's reference criteria to the Maryland data, resulting in three different sets of reference sites for comparison (Figure 2). The reference sites chosen from the Maryland data were most similar between the Virginia SCI and West Virginia SCI (excluding BPJ) methods, with the 150 Virginia-method reference sites forming a subset of the 209 West Virginia-method sites. This was a result of the reference criteria being the same, except that the Virginia method included total nitrogen and total phosphorus, as well as slightly stricter criteria for urban land use, bank disruptive pressure, and riparian buffer (Table 4). In contrast, 59 of the 144 reference sites (41%) selected using the Maryland B-IBI reference selection method were not selected using the Virginia or West Virginia methods. All but one of the 59 sites that met Maryland reference criteria—but were rejected using the Virginia or West Virginia methods—did not meet criteria related to

instream physical habitat (specifically total habitat score, channel alteration, bank stability, or bar formation). Conversely, sites that were selected as reference using the Virginia or West Virginia methods—but were rejected using the Maryland method—did not meet Maryland criteria for remoteness (66 sites), percent of forested land cover (23), or riparian width (15). The reference sites selected by all three methods met the same chemical reference criteria.

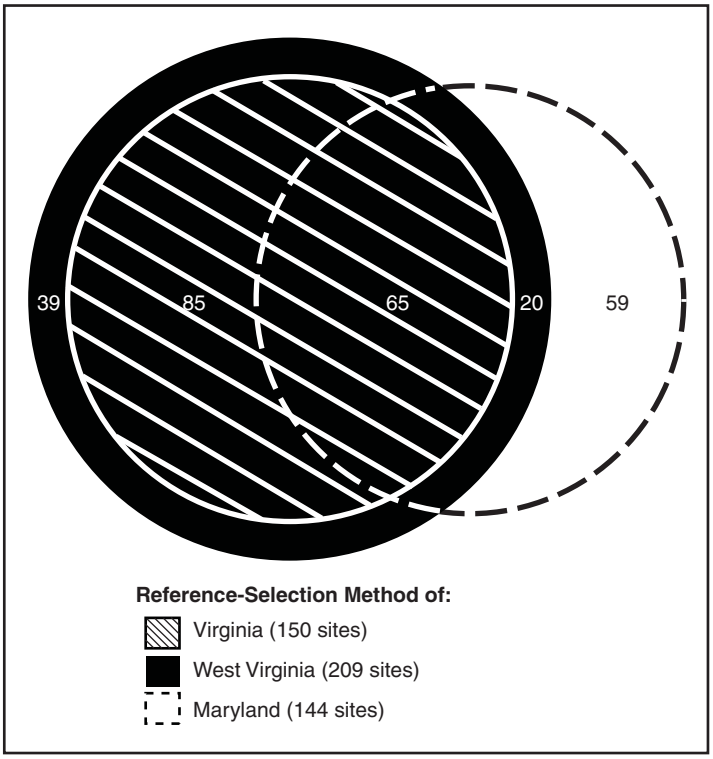

Figure 2. Venn diagram of reference sites selected from Maryland data using the Virginia, West Virginia, and Maryland reference criteria (described in Table 4). Numerals indicate the number of sites selected by all methods within the overlapping region.

Despite these differences in the specific reference sites selected by the Maryland method compared to the Virginia and West Virginia methods, the distributions of reference site SCI scores (using Maryland data) were similar for all three reference-selection methods (Figure 3). Nonparametric comparisons of the distributions of SCI scores for reference sites using the Kolmogorov-Smirnov test (Zar 1999) did not reveal significant differences among the three sets of reference sites using either the Virginia ($D = 1.21$, $P = 0.11$) or West Virginia ($D = 0.68$, $P = 0.75$) method for calculating SCIs. These data suggest that applying any of the three methods for selecting reference sites characterized the same range of stream conditions, even though different sites were selected (i.e., these different reference sites had the same stream quality). Note again that these analyses did not include the BPJ reference decisions of West Virginia.

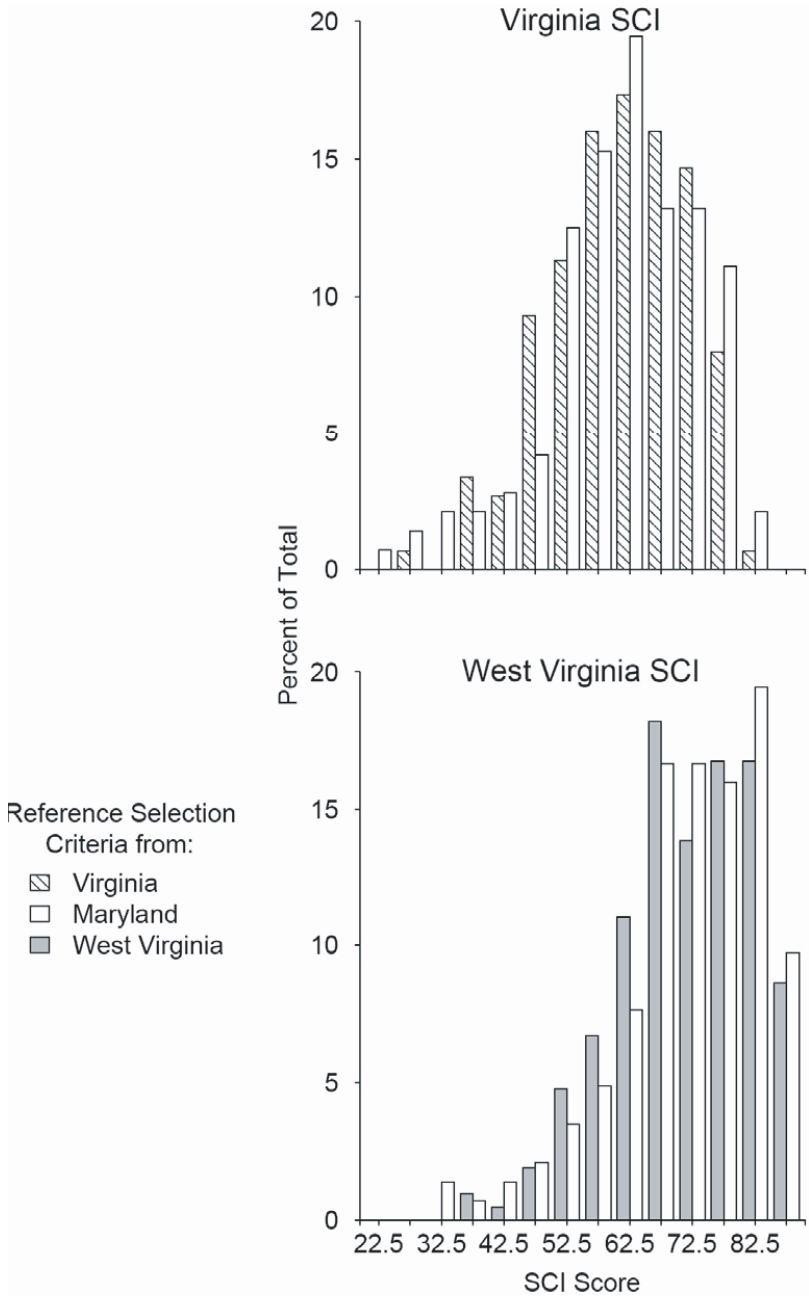

Figure 3. Distribution of SCI scores for reference sites selected from Maryland data using Virginia, West Virginia, and Maryland methods.

The SCI scores for each Maryland site were then used to generate cumulative distribution functions for the non-Coastal Plain region of Maryland (Figure 4). For each SCI score observed, we estimated the proportion of stream miles with that score by applying a 1 for streams with that score and a 0 for all other streams. The cumulative distribution was calculated by summing the proportion of streams with an equal or lower value for each SCI score. The standard error (SE) of all proportions was also estimated using stratified sampling estimators on the 1 or 0 scores for each stream, as described above.

The Virginia and West Virginia cumulative distribution curves of index scores on Maryland data were nearly identical (Figure 4) indicating that the SCI methods of these states characterized streams similarly. The correlation coefficient between Virginia and West Virginia SCI scores for Maryland data was $r = 0.96$. The correlation coefficient between West Virginia SCI and Maryland B-IBI scores for Maryland data was somewhat lower at $r = 0.86$. Using the degradation thresholds of the 10th and 5th percentiles of reference sites for Virginia and West Virginia, respectively, the two SCI methods rated streams similarly, suggesting that about 40% of Maryland streams were degraded. Although the Maryland B-IBI was not directly comparable to the SCIs, the proportion of streams rated as degraded and corresponding to 9th percentile of reference sites (i.e., below an B-IBI score of 3) was 53%, roughly corresponding with the proportion rated as degraded by the SCIs.

Because of the differences in the indicators and reference criteria used, the sampling results of the three states had to be adjusted before they could be combined into an integrated assessment of stream condition. The differences in indicators were resolved by calculating the West Virginia SCI for Maryland data (eliminating the Maryland B-IBI) so that Maryland results could be combined directly with the West Virginia SCI for West Virginia data and the Virginia SCI for Virginia data (which was very similar to the West Virginia SCI and could be used in its original form). An additional adjustment was needed to address the different reference criteria among the three states. Specifically, comparable reference criteria had to be applied to data from all three states and a threshold of degradation selected and its implications evaluated.

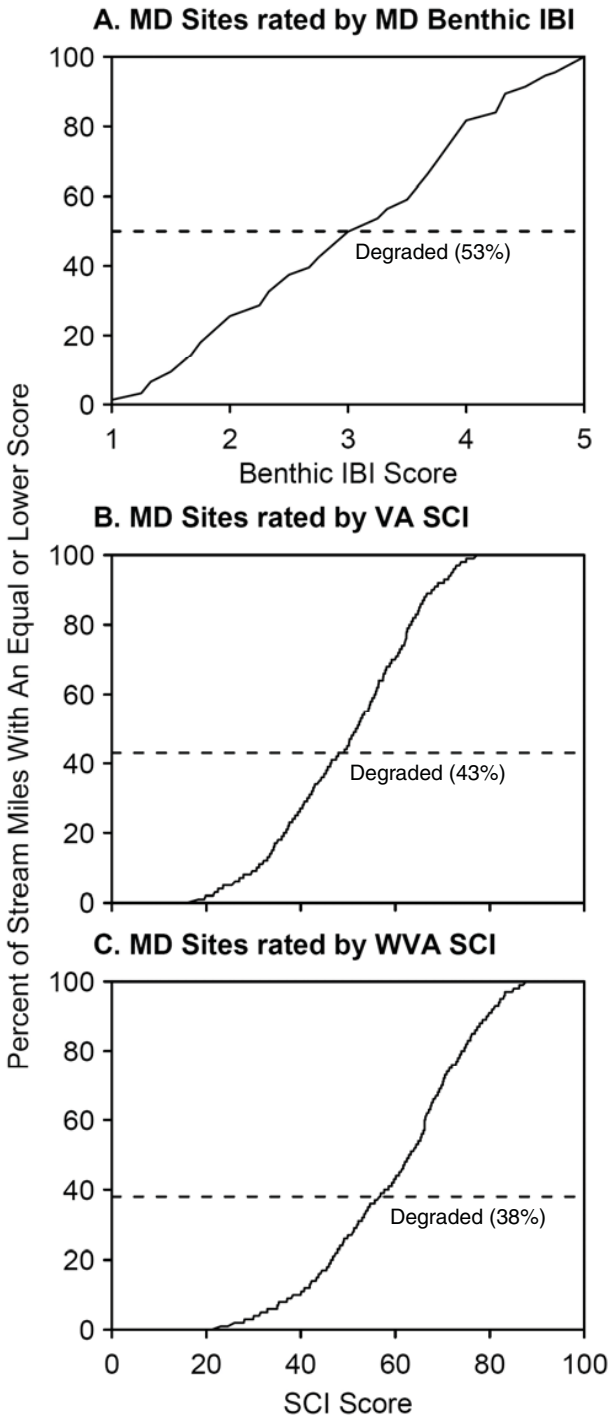

Figure 4. Cumulative distribution of Maryland benthic IBI scores (A), Virginia SCI scores (B), and West Virginia SCI scores (C) for Maryland streams randomly sampled during 2000-2003.

5. Integration of Assessments from the Three States

Given the comparability among the three state's methods (excluding the West Virginia BPJ reference decisions), we proceeded to integrate the results into a single estimate of stream condition using stratified sampling estimators that weighted each state estimate by the proportion of total stream miles contributed by each to the combined non-Coastal Plain region of the three states. The non-Coastal Plain stream miles for each state were as follows: 5,946 miles (7.2% of the three state total) in Maryland; 47,920 miles (58.2%) in Virginia; and 28,510 miles (34.6%) in West Virginia. Because the SCI methods produced such similar results, we compared the Virginia and West Virginia SCI distributions directly. For Maryland, we arbitrarily decided to apply the West Virginia SCI method to combine data. The combined cumulative distributions of scores from each state's sample of sites were calculated as described above (Figure 5).

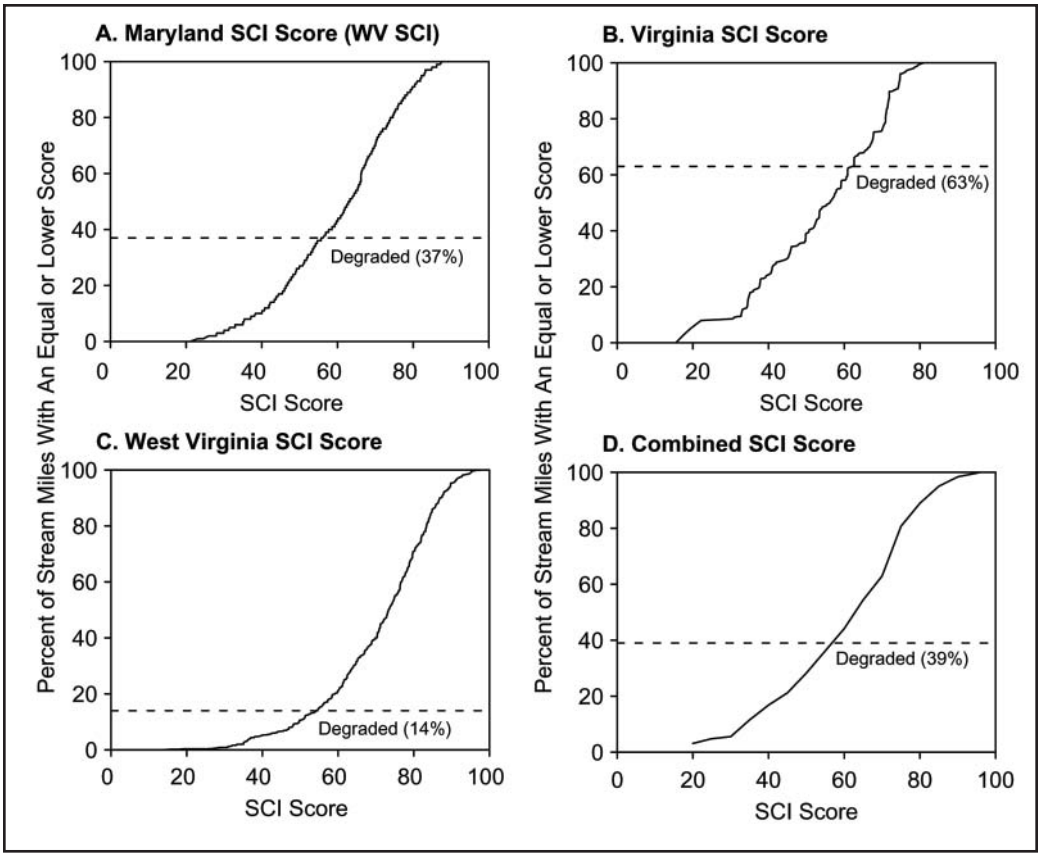

Figure 5. Cumulative distribution of SCI scores for Maryland (using WV SCI method), Virginia, West Virginia, and all three states combined. Thresholds for categorizing "degraded" condition correspond to the 10th percentile of the distribution of reference SCI scores.

Because the distributions of reference SCI scores were similar among states (Figure 3), we based our degradation thresholds for the assessment on percentiles taken from the combined set of reference sites from all three states (including reference sites that would have been eliminated by West Virginia BPJ). Using the 10th percentile of reference sites as the threshold of degradation, the cumulative distribution across states (Figure 5D) rated about 39% (SE= 4%) of the non-Coastal Plain streams in Maryland, Virginia, and West Virginia as degraded. Virginia had the greatest estimated proportion of degraded streams (63%, SE=8%; Figure 5) and West Virginia the least (14%, SE=1%). Maryland streams were intermediate, exhibiting an estimated 37% (SE=2%) degraded streams. Figure 5 illustrates how all three states can be combined using a single reference condition and threshold of degradation. Because West Virginia actually uses a stricter reference condition that includes site-by-site BPJ, the proportion of West Virginia streams rated as degraded are much higher when the reference sites including BPJ exclusions are used. Including BPJ reduced the number of reference sites from 349 to 216 and shifted the distribution of reference sites SCI scores toward higher values (Figure 6). This shift effectively raises the threshold that streams have to meet to be considered non-degraded. This analysis indicates that West Virginia has far fewer degraded streams than in Maryland or Virginia, a result that was not apparent using each state's independent assessments.

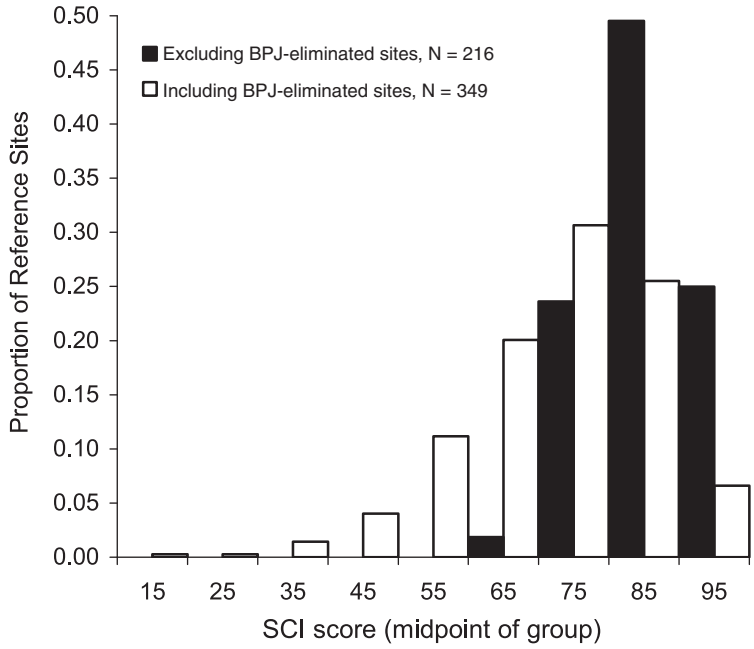

Figure 6. Distribution of SCI scores for West Virginia reference sites selected using all criteria including best professional judgment (BPJ) of anthropogenic disturbance (more restrictive; i.e., fewer sites), and all criteria except BPJ (less restrictive; i.e., more sites).

These higher degradation thresholds derived from the distribution of West Virginia reference sites selected using the additional BPJ decisions cannot be used in the combined non-Coastal Plain region of the three states, because these site-by-site decisions cannot be applied to the data from Maryland and Virginia. However, to illustrate the effect of using degradation thresholds based on higher quality reference sites (e.g., West Virginia's), we calculated the proportion of stream miles in the combined non-Coastal Plain region based on the 47th percentile of reference sites. This percentile was selected because the 47th percentile of West Virginia reference sites without BPJ exclusions corresponds to the 10th percentile of West Virginia reference sites using the BPJ exclusions. Therefore, the 47th percentile simulates the degradation threshold that would have been obtained if all three states used comparable BPJ exclusions. This higher threshold results in 64% of stream miles being rated as degraded (Figure 7). This difference in assessments indicates the importance of selecting a reference condition appropriate for water resource management goals.

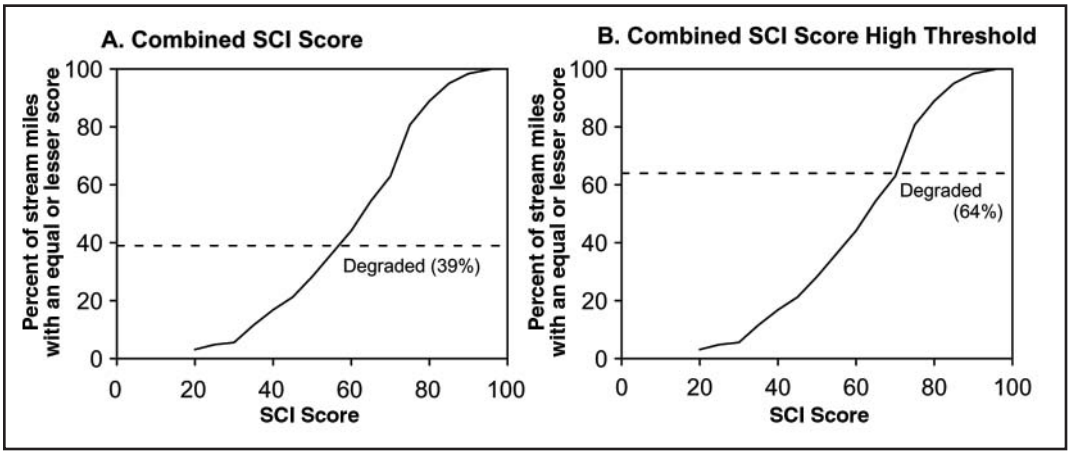

Figure 7. Cumulative distribution of SCI scores for all three states combined (Maryland using WV SCI method, Virginia, West Virginia) at two different thresholds for categorizing stream condition based on percentiles of reference sites. Left figure shows "degraded" condition corresponding to the 10th percentile of the distribution of reference SCI scores (no BPJ). Right figure corresponds to the 47th percentile of references scores (no BPJ) to simulate the effect that West Virginia BPJ would have had if applied to all states.

6. Discussion and Recommendations

This study demonstrates that state stream assessment programs can be integrated over larger regions if all states have similar sample frames, probabilistic survey designs, and comparable indicators (though differences in reference condition must be adjusted for). This assessment integration was possible even though differences in benthic macroinvertebrate collection procedures (related to sampling gear, habitats sampled, and level of taxonomic identification) precluded directly combining site data.

If the sample frames are different among the states, the integration must be restricted to the overlapping stream scale so that the population of interest is the same. In this study all three states used a 1:100,000-scale stream network. Differences in survey designs (if all designs are probability-based) do not affect integration, as each state is a de facto stratum that can be combined into an overall estimate.

The most challenging aspect of assessment integration is reconciling different indicators of stream condition. While Virginia and West Virginia used similar SCIs (metric scores based on the range of values at all sample sites and degradation thresholds applied as percentiles of reference SCI scores), Maryland used a conceptually different B-IBI (metric scores assigned relative to reference condition and averaged to produce B-IBI scores). We determined that the distribution of site scores (using Maryland data only) was very similar for the Virginia and West Virginia SCIs, but more uniform (i.e., a flatter line; see Figure 4) for the Maryland B-IBI, indicating the scores “stretched” more evenly across the full range of sites. We also noted that the wider distribution of site scores in the West Virginia data stretched the SCI scores relative to the Virginia SCIs as a result of their wider range of scores at all sites. These differences in indicators have resulted from independent indicator development in each state undertaken to address each state’s management objectives. Each indicator is based on sound principles and serves its state’s needs well. At the same time, such differences create problems for integration. While the B-IBI may perform better in Maryland, integration required that one of the SCIs be substituted for the B-IBI so that similar indicators could be used in all states. We calculated SCI scores (using the West Virginia method) on Maryland data for the final assessment integration.

In addition to calculating the same or similar indicators for all sites to be integrated, a single reference condition must be used to set assessment thresholds. Ideally, a single set of reference sites could be selected from all three states and used to develop and rate the indicators. Such a project, however, is time and resource intensive; another solution is to calibrate the reference conditions of each state on one set of sites (Maryland in this case). This requires a careful comparison of reference criteria and use of surrogate variables where necessary.

The West Virginia procedure for selecting reference sites included BPJ decisions that excluded additional sites with known human disturbances, and thus was more restrictive (i.e., fewer reference sites qualified). These BPJ decisions involved visual evaluations of candidate reference sites for signs of anthropogenic disturbances (e.g., surface mines) and nonpoint source pollution (e.g., livestock feedlots). In addition, BPJ was used to exclude sites with high conductivity and fecal coliform bacteria values when appropriate. These site-by-site decisions effectively set a higher standard of stream quality for West Virginia. These BPJ decisions could not be precisely defined (e.g., assigned standard values as is done with subjective habitat evaluations) and thus cannot (by definition) be replicated for other states. While including BPJ increases the confidence that West Virginia's reference sites are minimally disturbed (and may improve stream management), it is a barrier to integration.

After excluding West Virginia's BPJ, we applied each state's reference criteria to Maryland data and still produced different suites of reference sites. However, the distributions of scores from reference sites selected were similar, suggesting the different reference sites were of similar stream quality (comparably affected by human disturbance). Even though all reference criteria are incomplete surrogates of minimally disturbed condition, different reference criteria may be equally useful for selecting subsets of minimally disturbed streams.

In contrast, inclusion of BPJ dramatically affected the assessment of stream condition, as expected, increasing the proportion of degraded streams in non-Coastal Plain region of Maryland, Virginia, and West Virginia, from 39% to 64%. Again, this is a result of using higher quality reference sites that effectively raises the threshold as a higher percentile of the larger set of reference sites (i.e., from 10th to 47th). Therefore, it is critical that the proportion of degraded streams in the entire non-Coastal Plain region of Maryland, Virginia, and West Virginia, be determined using the same "yardstick" (i.e., similar SCIs with comparable reference conditions). Which yardstick is used depends on management objectives and the confidence that the reference sites are minimally disturbed.

This proof of concept study leads us to make the following recommendations for integrating stream assessment results among state programs:

- Sample frames must be comparable; if different map scales are used, potentially expensive geographic information system (GIS) analysis may be needed to determine the overlapping populations of streams.
- Different survey designs may be combined if they are probability-based, since individual states are strata in calculations of regional estimates.
- Results from different biological sampling procedures can be integrated if reference-based indicators are used to summarize the results.
- The ratings of stream condition will depend on how indicators are linked to reference condition, so a common reference condition ("yardstick") must be used to set thresholds of degradation.
- A common reference condition requires the application of objective criteria for which there are appropriate variables or surrogates for all sites, i.e., BPJ decisions that are not codified as standard values are not repeatable and cannot be used in integration.
- Both (1) modifications to state programs to make them more comparable and (2) the analyses to integrate results that are significantly different require staff and financial resources; therefore, we recommend that states collaborate early in their development of stream assessment programs to facilitate future integration.

7. Literature Cited

Barbour, M.T., Gerritsen, J., Snyder, B.D. and Stribling, J.B. 1999. "Rapid Bioassessment Protocols for using streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish." Second Edition. *EPA/841-B-99-002*. U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

Barbour, M.T., Stribling, J.B. and Karr, J.R. 1995. "Multimetric approach for establishing biocriteria and measuring biological condition." *In* W.S. Davis and T.P. Simon (eds.) *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making*. Lewis, Boca Raton, FL. Pages 63-77.

Berger, Y. 2004. "A simple variance estimator for unequal probability sampling without replacement." *Journal of Applied Statistics* 31:305-315.

Burton, J., and Gerritsen, J. 2003. *A stream condition index for Virginia non-Coastal Plain streams*. Prepared by Tetra Tech, Inc., for U.S. Environmental Protection Agency Office of Science and Technology, Office of Water, Washington, DC, U.S. EPA Region 3 Environmental Services Division, Wheeling, and WV Virginia Department of Environmental Quality, Richmond, VA. Available at www.deq.virginia.gov/watermonitoring/pdf/vastrmcon.pdf

Cochran, W.G. 1977. *Sampling Techniques*. Third ed. John Wiley and Sons, New York, NY.

Karr, J.R. 1991. "Biological integrity: A long-neglected aspect of water resource management." *Ecological Applications* 1:66–84.

Klauda, R., Kazyak, P., Stranko, S., Southerland, M., Roth, N. and Chaillou, J. 1998. "The Maryland Biological Stream Survey: A state agency program to assess the impact of anthropogenic stresses on stream habitat quality and biota." Third EMAP Symposium, Albany, NY. *Environmental Monitoring and Assessment* 51:299-316.

National Water Quality Monitoring Council (NWQMC). 2001. "Towards a Definition of Performance-Based Laboratory Methods." *National Water Quality Monitoring Council Technical Report 01–02*. U.S. Geological Survey, Reston, VA.

Paul, M.J., Stribling, J.B., Klauda, R., Kazyak, P., Southerland, M. and Roth, N. 2002. *A physical habitat index for freshwater wadeable streams in Maryland*. Prepared by Tetra Tech, Inc., Owings Mills, MD for Maryland Department of Natural Resources.

Paul, M.J., Stribling, J.B., Klauda, R., Kazyak, P., Southerland, M., and Roth, N. 2003. *Further Development of a Physical Habitat Index for Maryland Wadeable Freshwater Streams*. Report prepared by Versar, Inc., Columbia, MD; Tetra Tech, Inc., Owings Mills, MD; and Maryland Department of Natural Resources. CBWP-MANTA-EA-03-04.

Roth, N., Vølstad, J., Erb, L. and Weber, E. 2005. *Maryland Biological Stream Survey 2000-2004 Volume 6: Laboratory, Field, and Analytical Methods*. DNR 12-0305-0108. Maryland Department of Natural Resources, Monitoring and Non-tidal Assessment Division, Annapolis, MD.

Roth, N.E., Vølstad, J.H., Mercurio, G., and Southerland, M.T. 2002. *Biological Indicator Variability and Stream Monitoring Program Integration: A Maryland Case Study*. Prepared by Versar, Inc., Columbia, MD, for U.S. Environmental Protection Agency, Office of Environmental Information and the Mid-Atlantic Integrated Assessment Program.

Southerland, M., Rogers, G., Kline, M., Morgan, R., Boward, D., Kazyak, P. and Stranko, S. 2005. *New Biological Indicators to Better Assess Maryland Streams* (DRAFT). Prepared for Monitoring and Non-Tidal Assessment Division, Maryland Department of Natural Resources, Annapolis, MD.

Stevens, D.L., Jr. 1997. "Variable density grid-based sampling designs for continuous spatial populations." *Environmetrics* 8:167-95.

Stevens, D.L., Jr. and Olsen, A.R. 2004. "Spatially balanced sampling of natural resources." Theory and Methods. *Journal of American Statistical Association* 99:262-278.

Strahler, A.N. 1957. "Quantitative analysis of watershed geomorphology." *Transactions of the American Geophysical Union* 38:913-920.

Tetra Tech, Inc. 2000. *A stream condition index for West Virginia wadeable streams*. Prepared for U.S. EPA Region 3 Environmental Services Division, and U.S. EPA Office of Science and Technology, Office of Water. Available at www.dep.state.wv.us//show_blob.cfm?id=536&name=WV-Index.pdf

WVDEP. 2006. Standard Operating Procedures. Watershed Branch. Jeffrey Bailey, primary author (Working Document).

U.S. Environmental Protection Agency (EPA). 2000. Guidance for the Data Quality Objectives Process (QA/G-4). *EPA/600/R-96/055*. Washington, DC.

U.S. EPA. 2002a. Research strategy, Environmental monitoring and assessment program. *EPA 620-R/02-2002.* Research Triangle Park, NC.

U.S. EPA. 2002b. "Summary of Biological Assessment Programs and Biocriteria Development for States, Tribes, Territories, and Interstate Commissions: Streams and Wadeable Rivers." *EPA 822-R-02-048*. Washington, DC.

Virginia Department of Environmental Quality (VDEQ). 2003. "The quality of Virginia non-tidal streams: first year report." *VDEQ Technical Bulletin WQA/2002-001*. Richmond, VA. Available at http://www.deq.virginia.gov/water/probmon.pdf.

VDEQ. 2005. "Using Probabilistic Monitoring Data to Validate the Virginia Stream Condition Index (DRAFT)." VDEQ Technical Bulletin WQA/2005-002. Richmond, VA. Available at http://www.deq.state.va.us/probmon/pdf/scival.pdf

Vølstad, J.H., Roth, N.E., Southerland, M.T. and Mercurio, G. 2003. "Pilot Study for Montgomery County and Maryland DNR Data Integration: Comparison of Benthic Macroinvertebrate Sampling Protocols for Freshwater Streams." *EPA/903/R-03/005*. U.S. Environmental Protection Agency Region 3, Office of Environmental Information and Mid-Atlantic Integrated Assessment Program, Fort Meade, MD.

West Virginia Department of Environmental Protection (WVDEP). 2005. *West Virginia's enhanced water quality monitoring strategy.* Prepared by Watershed Branch, Division of Water and Waste Management. Available at www.dep.state.wv.us/Docs/8949_W.Va._Enhanced_Monitoring_Strategy.pdf

WVDEP. 2006. Standard Operating Procedures. Watershed Branch. Jeffrey Bailey, primary author (Working Document)

Zar, J.H. 1999. *Biostatistical Analysis*. 4th Edition. Prentice-Hall, Inc., Upper Saddle River, NJ.

United States
Environmental Protection
Agency

Office of Environmental Information
Mid-Atlantic Integrated Assessment
Environmental Science Center
701 Mapes Road
Fort Meade, MD 20755-5350

Please make all necessary changes on the below label, detach or copy, and return to the address in the upper left-hand corner.

If you do not wish to receive these reports, CHECK HERE ☐; detach, or copy this cover, and return to the address in the upper left-hand corner.

PRESORTED STANDARD
POSTAGE & FEES PAID
EPA
PERMIT No. G-35

Official Business
Penalty for Private Use
$300

EPA/903/R-05/003
June 2006