BUREAU OF POINT AND NON-POINT SOURCE MANAGEMENT

AN INDEX OF BIOTIC INTEGRITY FOR BENTHIC MACROINVERTEBRATE COMMUNITIES IN PENNSYLVANIA'S WADEABLE, FREESTONE, RIFFLE-RUN STREAMS

JULY 2013

Prepared by:

Brian Chalfant
PA Department of Environmental Protection
Bureau of Point and Non-Point Source Management
11th Floor: Rachel Carson State Office Building
Harrisburg, PA 17105

INTRODUCTION

The Pennsylvania Department of Environmental Protection (PADEP) developed an index of biotic integrity (IBI) for benthic macroinvertebrate communities in Pennsylvania's wadeable, freestone, riffle-run streams. Through direct quantification of biological attributes along a gradient of ecosystem conditions, this IBI measures the extent to which anthropogenic activities compromise a stream's ability to support healthy aquatic communities (Davis and Simon 1995). This biological assessment tool helps guide and evaluate aquatic resource legislation, policy, goals, and management strategies (Davis and Simon 1995; Davies and Jackson 2006; Hawkins 2006).

BIOLOGICAL SAMPLING METHODS

This IBI applies to benthic macroinvertebrate samples collected from wadeable, freestone, riffle-run streams in Pennsylvania using a D-frame net with 500-micron mesh. Field sampling and laboratory methods are more fully described in PADEP's Standardized Biological Field Collection and Laboratory Methods, Section V (PADEP 2003). At a sampling site, biologists work progressively upstream, compositing six kicks from riffle areas distributed throughout a 100-meter stream reach. Biologists aim to sample areas representative of the variety of riffle habitats (e.g., slower flowing, shallow riffles and faster flowing, deeper riffles) present in the sample reach. With each kick, biologists aim to disturb approximately one square meter immediately upstream of the net for approximately one minute to an approximate depth of 10 cm, as substrate allows. Composited samples are preserved with 95% ethanol in the field and transported back to the laboratory for processing.

In the lab, each composited sample is placed into a 3.5" deep rectangular pan (measuring 14" long x 8" wide on the bottom of the pan) marked off into 28 four-square inch (2" x 2") grids. Four of the grids are randomly selected. The contents of the selected grids are extracted from within four-square inch circular "cookie cutters" placed in the randomly selected grids in the pan, using plastic spoons, knives, turkey basters, and other implements as needed. These extracted contents are then placed into a second pan with the same dimensions and markings as the initial pan. All the organisms are picked from this second pan.

If less than 160 identifiable organisms are picked from the second pan, an additional grid is randomly selected and extracted from the first pan. The contents of this additional grid are transferred to the second pan, and the organisms are picked from the second pan. This process is continued until the target number of organisms is reached. The target number of organisms is 200 ± 40 identifiable organisms, with 190 to 210 identifiable organisms being the preferred range. In situations with a count of identifiable organisms in a sub-sample between 160 and 180 and a sample that has not been entirely picked, PADEP highly encourages picking an additional grid or two to get closer to the target number of 200 identifiable organisms (i.e., in the preferred 190 to

210 organism range) as long as picking additional grids will not result in a sub-sample with more than 240 identifiable organisms.

If more than 240 identifiable organisms are picked from the initial four grids, then those organisms are all placed into another pan and floated. A grid is then randomly selected and the organisms are picked from the selected grid. This process continues until the target number of organisms (200 ± 40, with 190 to 210 preferred) is reached.

Any grid selected during any part of the sub-sampling process is picked in its entirety. The total number of grids selected for each part of the sub-sampling process (e.g., 4 of 28 grids from the first pan, 10 of 28 grids from the second pan) is recorded.

Organisms in the sub-sample are identified and counted. Midges are identified to the family level of Chironomidae. Snails, clams, and mussels are all also identified to family levels. Roundworms and proboscis worms are identified to the phylum levels of Nematoda and Nemertea, respectively. Moss animacules are identified to the phylum level of Bryozoa. Flatworms and leeches are identified to the class levels of Turbellaria and Hirudenia, respectively. Segmented worms, aquatic earthworms, and tubificids are identified to the class level of Oligochaeta. All water mites are identified as Hydracarina, an artificial taxonomic grouping of several mite superfamilies. All other macroinvertebrates are identified to genus level.

THE METRICS

A number of different metric combinations were evaluated during index development. The following six metrics were selected for inclusion in the IBI based on various performance characteristics. These six metrics all exhibited a strong ability to distinguish between relatively pristine and heavily impacted conditions. In addition, these six metrics measure different aspects of the benthic macroinvertebrate communities. When used together in a multimetric index, these six metrics provide a solid foundation for assessing the biological condition of benthic macroinvertebrate assemblages in Pennsylvania's wadeable, freestone, riffle-run stream ecosystems.

Total Taxa Richness

This taxonomic richness metric is a count of the total number of taxa in a sub-sample. Generally, this metric is expected to decrease with increasing anthropogenic stress to a stream ecosystem, reflecting loss of taxa and increasing dominance of a few pollution-tolerant taxa. Other benefits of including this metric include its common use in many biological monitoring and assessment programs in other parts of the world as well as its ease of explanation and calculation.

Ephemeroptera + Plecoptera + Trichoptera Taxa Richness
(Pollution Tolerance Values 0-4 only)

This taxonomic richness metric is a count of the number of taxa belonging to the orders Ephemeroptera, Plecoptera, and Trichoptera (EPT) in a sub-sample – common names for these orders are mayflies, stoneflies, and caddisflies, respectively. The aquatic life stages of these three insect orders are generally considered sensitive to, or intolerant of, many types of pollution (Lenat and Penrose 1996), although sensitivity to different types of pollution varies among taxa in these insect orders. The version of this metric used here only counts EPT taxa with PTVs of 0 to 4, excluding a few of the most tolerant mayfly and caddisfly taxa. This metric is expected to decrease in value with increasing anthropogenic stress to a stream ecosystem, reflecting the loss of taxa from these largely pollution-sensitive orders. This metric has a history of use across the world and is relatively easy to use, explain, and calculate (Lenat and Penrose 1996).

Beck's Index (version 3)

This taxonomic richness and tolerance metric is a weighted count of taxa with pollution tolerance values of 0, 1, or 2. The name and conceptual basis of this metric are derived from the water quality work of William H. Beck in Florida (Beck 1955). This metric is expected to decrease in value with increasing anthropogenic stress to a stream ecosystem, reflecting the loss of pollution-sensitive taxa. It should be noted that the version of the Beck's Index metric used for this project, although similar in name and concept, differs slightly in its calculation from the Beck's Index used in PADEP's multihabitat protocol for assessing biological condition of low gradient, pool-glide type streams.

Shannon Diversity

This community composition metric measures taxonomic richness and evenness of individuals across taxa of a sub-sample. This metric is expected to decrease in value with increasing anthropogenic stress to a stream ecosystem, reflecting loss of pollution-sensitive taxa and increasing dominance of a few pollution-tolerant taxa. The name and conceptual basis for this metric are derived from the information theory work of Claude Elwood Shannon (Shannon 1968).

Hilsenhoff Biotic Index

This community composition and tolerance metric is calculated as an average of the number of individuals in a sub-sample, weighted by pollution tolerance values. Developed by William Hilsenhoff, the Hilsenhoff Biotic Index (Hilsenhoff 1977, 1987, 1988; Klemm et al. 1990) generally increases with increasing ecosystem stress, reflecting increasing dominance of pollution-tolerant organisms.

Percent Sensitive Individuals
(Pollution Tolerance Values 0-3 only)

This community composition and tolerance metric is the percentage of individuals with pollution tolerance values of 0 to 3 in a sub-sample and is expected to decrease in value with increasing anthropogenic stress to a stream ecosystem, reflecting loss of pollution-sensitive organisms.

Example calculations for each metric are provided below for a sub-sample from Lycoming Creek in Lycoming County collected November 19, 2001.

Benthic macroinvertebrate sub-sample from Lycoming Creek in Lycoming County collected November 19, 2001

Taxa Name	Number of Individuals	Pollution Tolerance Value
Acentrella	1	4
Isonychia	4	3
Epeorus	6	0
Leucrocuta	1	1
Rhithrogena	9	0
Stenonema	8	3
Ephemerella	32	1
Serratella	1	2
Paraleptophlebia	4	1
Pteronarcys	1	0
Taeniopteryx	1	2
Leuctra	2	0
Agnetina	1	2
Paragnetina	1	1
Chimarra	1	4
Dolophilodes	1	0
Cheumatopsyche	25	6
Hydropsyche	22	5
Rhyacophila	16	1
Glossosoma	2	0
Brachycentrus	3	1
Micrasema	1	2
Apatania	2	3
Psilotreta	1	0
Psephenus	3	4
Optioservus	7	4
Atherix	1	2
Antocha	2	3
Hexatoma	5	2
Prosimulium	1	2
Chironomidae	49	6
Ancylidae	2	7
Oligochaeta	1	10

Total Taxa Richness

There are ***33 taxa*** in this sub-sample, so

Total Taxa Richness = 33

EPT Taxa Richness (PTV 0-4 only)

There are ***9 Ephemeroptera taxa*** (Acentrella, Isonychia, Epeorus, Leucrocuta, Rhithrogena, Stenonema, Ephemerella, Serratella, Paraleptophlebia), ***5 Plecoptera taxa*** (Pteronarcys, Taeniopteryx, Leuctra, Agnetina, Paragnetina) and ***8 Trichoptera taxa*** (Chimarra, Dolophilodes, Rhyacophila, Glossosoma, Brachycentrus, Micrasema, Apatania, Psilotreta) in this sub-sample ***with PTVs $\leq$ 4***, so

EPT Taxa Richness (PTV 0-4 only) = 9 + 5 + 8

EPT Taxa Richness (PTV 0-4 only) **= 22**

Beck's Index (version 3)

Beck's Index (version 3) =
(3 x (number of taxa with PTV = 0)) +
(2 x (number of taxa with PTV = 1)) +
(1 x (number of taxa with PTV = 2))

There are ***7 taxa in this sub-sample with PTV = 0***. There are ***6 taxa in this sub-sample with PTV = 1***. There are ***7 taxa in this sub-sample with PTV = 2***, so

Beck's Index (version 3) = 3(7) + 2(6) + 1(7)

Beck's Index (version 3) = 21 + 12 + 7

Beck's Index (version 3) **= 40**

Hilsenhoff Biotic Index

$$\text{Hilsenhoff Biotic Index} = \sum_{i=0}^{10} [(i * n_{indvPTVi})] / N$$

where $n_{indvPTVi}$ = the number of individuals in a sub-sample with PTV of i and N = the total number of individuals in a sub-sample

In this sub-sample,

there are 22 individuals with PTV = 0,	there are 22 individuals with PTV = 5
there are 57 individuals with PTV = 1,	there are 74 individuals with PTV = 6
there are 11 individuals with PTV = 2,	there are 2 individuals with PTV = 7
there are 16 individuals with PTV = 3,	there are 0 individuals with PTV = 8 or 9, and
there are 12 individuals with PTV = 4,	there is 1 individual with PTV = 10.

There are a total of 217 individuals in this sub-sample, so

$$\text{Hilsenhoff Biotic Index} = [(0 * 22) + (1 * 57) + (2 * 11) + (3 * 16) + (4 * 12) + (5 * 22) + (6 * 74) + (7 * 2) + (8 * 0) + (9 * 0) + (10 * 1)] / 217$$

Hilsenhoff Biotic Index = 3.47

Shannon Diversity Index

$$\text{Shannon Diversity Index} = -\sum_{i=1}^{Rich} (n_i / N) \ln (n_i / N)$$

where n_i = the number of individuals in each taxon (relative abundance); N = the total number of individuals in a sub-sample; and Rich = the total number of taxa in a sub-sample (total taxa richness)

There are 33 taxa in this sub-sample. The numbers of individuals in each taxon are shown in the table above. There are a total of 217 individuals in the sub-sample, so

$$\text{Shannon Diversity Index} = -1[(1 / 217) \ln (1 / 217) + (4 / 217) \ln (4 / 217) + (6 / 217) \ln (6 / 217) + (1 / 217) \ln (1 / 217) + (9 / 217) \ln (9 / 217) + (8 / 217) \ln (8 / 217) + (32 / 217) \ln (32 / 217) + (1 / 217) \ln (1 / 217) + \ldots \textit{(do this for all 33 taxa)} \ldots (1 / 217) \ln (1 / 217)]$$

Shannon Diversity Index = 2.67

Percent Sensitive Individuals (PTV 0-3 only)

Percent Sensitive Individuals (PTV 0-3 only) = $\left(\sum_{i=0}^{3} n_{indvPTVi}\right) / N * 100$

where $n_{indvPTVi}$ = the number of individuals in a sub-sample with PTV of i and N = the total number of individuals in a sub-sample

In this sub-sample,

there are 22 individuals with PTV = 0,
there are 57 individuals with PTV = 1,
there are 11 individuals with PTV = 2, and
there are 16 individuals with PTV = 3.

There are a total of 217 individuals in this sub-sample, so

Percent Sensitive Individuals (PTV 0-3 only) = (22 + 57 + 11 + 16) / 217 *100

Percent Sensitive Individuals (PTV 0-3 only) = 106 / 217 * 100

Percent Sensitive Individuals (PTV 0-3 only) **= 48.8%**

THE INDEX

An index is simply a means to integrate information from various metrics of biological integrity (Barbour et al. 1999). In order to compare and combine sundry measures (e.g., percentage of individuals, counts of taxa, unitless numbers) of biological condition in a meaningful manner, it is necessary to standardize metrics with some mathematical transformation that results in a logical progression of values (Barbour et al. 1995).

To account for natural changes in benthic biota with stream size, different metric standardization values for samples from larger streams and smaller streams were developed for this IBI. Data suggest that the small stream approach is usually appropriate for first, second, and third order streams (using the Strahler stream ordering system) draining less than 25 to 50 square miles, while the large stream approach is usually appropriate for fifth order and larger streams draining more than 50 square miles. More detailed guidelines for deciding whether to apply the large-stream or small-stream metric standardization values to a sample are discussed below.

The one selected core metric that increases in value with increasing anthropogenic stress – Hilsenhoff Biotic Index – was standardized to approximately the 5th percentile of metric scores for all samples from smaller streams and for all samples from larger streams in the IBI development dataset to arrive at the respective small-stream and large-stream standardization values. Core metrics that decrease in value with increasing stress – Total Taxa Richness, EPT Taxa Richness, Beck's Index, Shannon Diversity, and Percent Sensitive Individuals – were standardized to approximately the 95th percentile of metrics scores for all samples from smaller streams and for all samples from larger streams in the IBI development dataset to set the respective small-

stream and large-stream standardization values. The following table presents small-stream and large-stream standardization values used for each core metric.

Metric	Metric Standardization Values	
	smaller streams most 1st to 3rd order < 25 square miles	**larger streams** most 5th order and larger > 50 square miles
Total Taxa Richness	33	31
EPT Taxa Richness (PTV 0-4 only)	19	16
Beck's Index (version 3)	38	22
Hilsenhoff Biotic Index	1.89	3.05
Shannon Diversity	2.86	2.86
Percent Sensitive Individuals (PTV 0-3 only)	84.5	66.7

To calculate the index of biological integrity, observed metric values are first standardized using the standardization values shown in the table immediately above and the following standardization equations.

The Hilsenhoff Biotic Index metric values are expected to increase in value with increasing anthropogenic stress and are standardized using the following equation:

Hilsenhoff Biotic Index standardized score =
(10 – observed value) / (10 – standardization value) * 100

The other five core metrics values are expected to decrease in value with increasing anthropogenic stress and are standardized using the following equation:

Standardized metric score = observed value / standardization value * 100

Once the observed metric values are standardized, the standardized metric scores are adjusted to maximum value of 100 if necessary. By standardizing metrics and setting a maximum value of 100 for the standardized metrics, the resulting adjusted standardized metric scores can range from maximum values of 100 to minimum values of zero, with scores closer to zero corresponding to increasing deviation from the expected reference condition and progressively higher values corresponding more closely to the biological reference condition (Barbour et al. 1995). This approach establishes upper bounds on the expected condition and moderate effects of metrics that may respond in some manner other than a monotonic response to stress. The index of biological integrity is calculated by calculating the arithmetic mean of these adjusted standardized metric values for the six core metrics, resulting in a multimetric index of biological integrity score that can range from 0 to 100. To get a score of zero, a sample would have to contain no organisms at all.

In order to incorporate the variability of metric scores with annual seasons in setting biological expectations, PADEP chose to implement different use attainment benchmarks as discussed below rather than adjust metric standardization values.

The sample from Lycoming Creek presented above was collected from a fifth order site draining approximately 173 square miles of land, so we will apply the large-stream

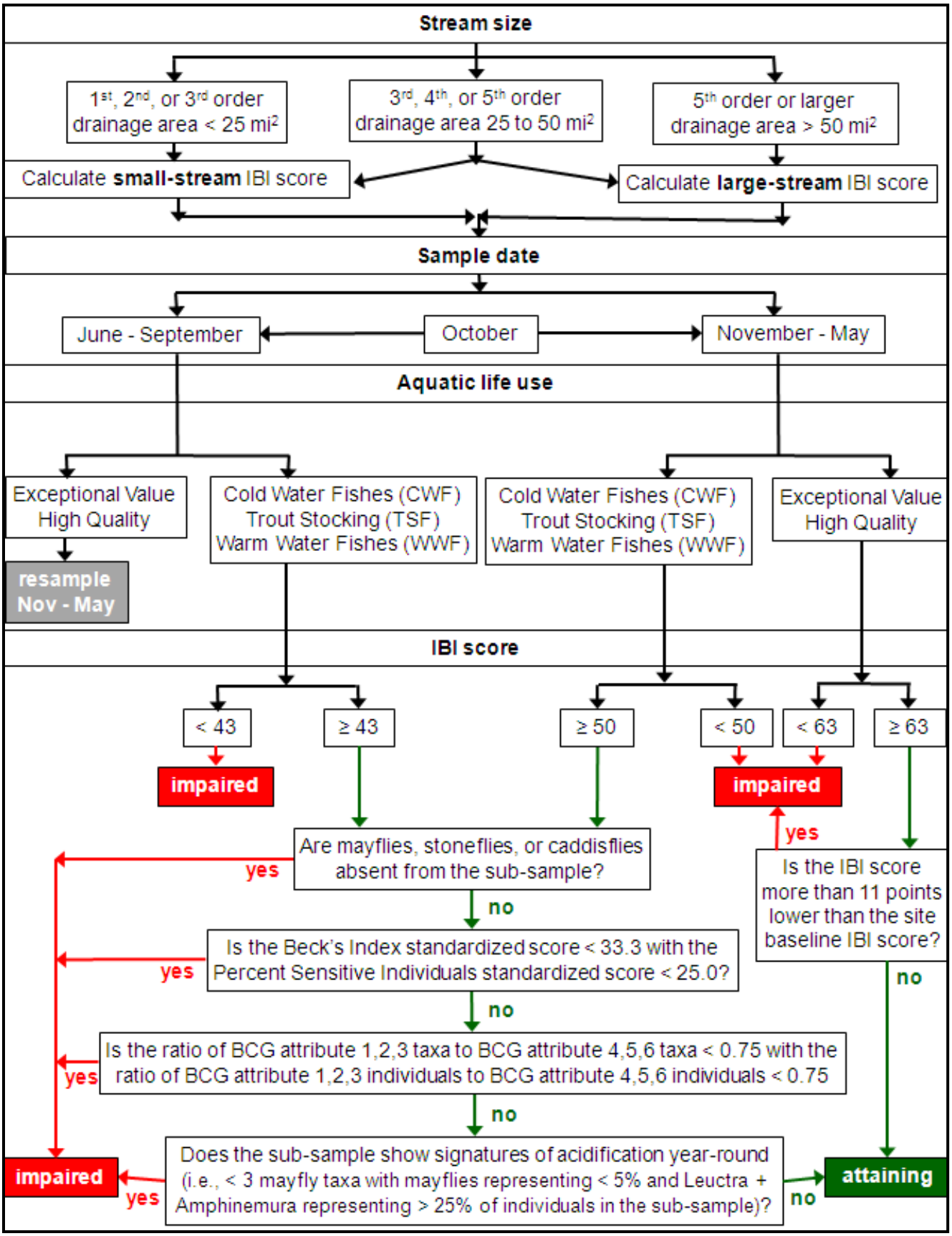
metric standardization values in the example metric standardization and index calculations presented in the table below. For a small-stream sample, we would simply substitute the small-stream metric standardization values in place of the large-stream metric standardization values – the rest of the index calculation process is the same regardless of stream size.

Metric	**Standardization Equation** (using large-stream standardization values)	**Observed Metric Value**	**Standardized Metric Score**	**Adjusted Standardized Metric Score** Maximum = 100
Total Taxa Richness	(observed value / 31) * 100	33	106.5	100
EPT Taxa Richness (PTV 0-4 only)	(observed value / 16) * 100	22	137.5	100
Beck's Index (version 3)	(observed value / 22) * 100	40	181.8	100
Hilsenhoff Biotic Index	[(10 – observed value) / (10 – 3.05)] * 100	3.47	94.0	94.0
Shannon Diversity	(observed value / 2.86) * 100	2.67	93.4	93.4
Percent Sensitive Individuals (PTV 0-3 only)	(observed value / 66.7) * 100	48.8	73.2	73.2
Arithmetic mean of adjusted standardized core metric scores = **IBI Score =**				**93.4**

AQUATIC LIFE USE ATTAINMENT BENCHMARKS

Due to the influences of annual seasons and drainage area seen in the dataset, PADEP recognizes different assessment tools and use attainment thresholds are appropriate for samples collected during different times of the year and from different size stream systems. It is noted that some site-specific exceptions to any thresholds may exist because of local scale natural limitations (e.g., habitat availability) on biological condition (Hughes 1995).

Based on the results of technical analyses, professional workshops, feedback from PADEP biologists and other colleagues, as well as policy considerations, PADEP implements a multi-tiered benchmark decision process for wadeable, freestone, riffle-run streams in Pennsylvania that incorporates stream size and sampling season as factors for determining aquatic life use attainment and impairment based on benthic macroinvertebrate sampling. A simplified flowchart of this decision process is outlined in the diagram below. Although this simplified decision matrix should guide most assessment decisions for benthic macroinvertebrate samples from Pennsylvania's wadeable, freestone, riffe-run streams using the collection and processing methods discussed above, situations exist where this simplified assessment schematic will not apply exactly as outlined – some such situations are discussed in the following text.


Stream size
1st, 2nd, or 3rd order drainage area < 25 mi²
3rd, 4th, or 5th order drainage area 25 to 50 mi²
5th order or larger drainage area > 50 mi²
Calculate small-stream IBI score
Calculate large-stream IBI score
Sample date
June - September
October
November - May
Aquatic life use
Exceptional Value High Quality
Cold Water Fishes (CWF) Trout Stocking (TSF) Warm Water Fishes (WWF)
Cold Water Fishes (CWF) Trout Stocking (TSF) Warm Water Fishes (WWF)
Exceptional Value High Quality
resample Nov - May
IBI score
< 43
≥ 43
≥ 50
< 50
< 63
≥ 63
impaired
impaired
yes
Are mayflies, stoneflies, or caddisflies absent from the sub-sample?
yes
no
Is the IBI score more than 11 points lower than the site baseline IBI score?
Is the Beck's Index standardized score < 33.3 with the Percent Sensitive Individuals standardized score < 25.0?
yes
no
no
Is the ratio of BCG attribute 1,2,3 taxa to BCG attribute 4,5,6 taxa < 0.75 with the ratio of BCG attribute 1,2,3 individuals to BCG attribute 4,5,6 individuals < 0.75
yes
no
Does the sub-sample show signatures of acidification year-round (i.e., < 3 mayfly taxa with mayflies representing < 5% and Leuctra + Amphinemura representing > 25% of individuals in the sub-sample)?
impaired
yes
no
attaining

The first step in the aquatic life use assessment process for wadeable, freestone, riffle-run streams in Pennsylvania based on benthic macroinvertebrate sampling considers stream size. PADEP does not feel that it is appropriate to set a single cutoff drainage area or stream order threshold to define which set of metric standardization values and which resulting IBI (i.e., large-stream or small-stream) should be applied. However – as stated above – data suggest that the small-stream approach is usually appropriate for samples from first, second, and third order streams draining less than 25 square miles of land, while the large-stream approach is usually appropriate for samples from fifth order and larger streams draining more than 50 square miles.

There are many important considerations when deciding whether to apply the small-stream or large-stream metric standardization values to a sample. Many stream systems experience a variety of changes as they flow from headwaters on downstream. These changes include, but are certainly not limited to changes in canopy shading, energy dynamics, algal growth, erosional and depositional patterns, habitat distributions, water temperature, and flow regimes. These shifts manifest themselves uniquely in each watershed. Streams in more northern, high elevation, high relief areas of the state may maintain cooler water, flashier flows, larger-particle substrates, and other characteristics typical of smaller streams at comparable drainage areas or stream orders when compared with streams in more southern, low elevation, low relief areas of the state. Local climatological and geological patterns also affect a stream's character.

When deciding which set of metric standardization values (i.e., small-stream or large-stream) to apply, care should be taken not to conflate human-induced changes to streams with natural landscape and climatological variations. For example, a stream draining 26 square miles of mostly corn and soybean fields with little forested riparian buffer may experience warmer water temperatures and more silted substrates than a stream of similar size draining a more forested watershed. The warmer water and more silted substrates of the agricultural stream may be characteristics typical of larger streams, but if those characteristics are primarily human-induced, then that argues against applying the large-stream metric standardization values based on the presence of those characteristics in the stream.

For streams of intermediate size (i.e., third, fourth, and some fifth order streams draining between 25 and 50 square miles of land), it will often be informative to consider both the small-stream and large-stream IBI scores and associated benchmarks. For example, if a sample from a fourth order site draining 30 square miles scores 77.0 on the small-stream IBI and 90.2 on the large-stream IBI and passes the additional screening questions, both approaches indicate aquatic life use attainment, so the use assessment decision is the same regardless of which set of metric standardization values is applied. In another instance, a sample collected in mid-March from a site draining 36 square miles may score 44.1 on the small-stream IBI – indicating impairment – while scoring 51.2 on the large-stream IBI – indicating possible attainment. Here, the small-stream and large-stream IBI score assessment decisions diverge. In such situations it may be especially useful to consider the additional screening questions – detailed below – when making an assessment decision.

The second step in the aquatic life use assessment process for wadeable, freestone, riffle-run streams in Pennsylvania based on benthic macroinvertebrate sampling considers sampling season. Samples collected during summer and early autumn months (i.e. June through September) are held to different IBI attainment thresholds than samples collected November through May since benthic macroinvertebrate communities in most wadeable, freestone, riffle-run streams in Pennsylvania exhibit consistent patterns of lower taxonomic diversity and organismal abundance during the summer and early autumn months compared with other times of the year. These seasonal index periods are intended as general guidelines and may vary slightly year-to-year depending on local climatological conditions. For example, a sample collected from a low elevation, low latitude stream during the last week of May in a particularly hot, dry year may be more properly evaluated using procedures set forth for the summer months – especially if many mayflies have already emerged from the stream – while a sample collected from a high elevation, high latitude location during the first week of June in an uncharacteristically cool, wet year may be more properly evaluated using the November to May procedures – especially if many mayfly nymphs are still present in the benthos.

October often is a transitional time for benthic macroinvertebrate communities in Pennsylvania with samples from earlier in the month resembling late summer communities (e.g., relatively low diversity and abundance) and samples from later in the month resembling early winter communities (e.g., increasing abundance of winter stoneflies). Therefore, depending on local climate, basin geology, and other factors discussed above (e.g., latitude, elevation, basin relief) samples from October may be evaluated using the June to September benchmarks or the November to May benchmarks. PADEP advises against sampling in mid-October to avoid these issues.

For samples collected between November and May, IBI scores < 50 result in aquatic life use impairment. Samples collected during these months scoring ≥ 50 on the appropriate IBI are subject to four screening questions before the aquatic life use can be considered attaining. **These additional screening questions are:**

1. **Are mayflies, stoneflies, or caddisflies absent from the sub-sample?** Organisms representing these three taxonomic orders are usually found in most healthy wadeable, freestone, riffle-run streams in Pennsylvania. If any or all of these orders are absent from a sample, this strongly suggests some sort of anthropogenic impact. Samples where one of these taxonomic orders is absent due to natural conditions (e.g., mayflies absent from a low-pH tannic stream) should be evaluated accordingly.

2. **Is the standardized metric score for the Beck's Index metric < 33.3 with the standardized metric score for the Percent Sensitive Individuals metric < 25.0?** Although these two metrics go into the IBI calculations, this screening question serves to double check that a sample has substantial richness and abundance of the most sensitive organisms.

3. **Is the ratio of Biological Condition Gradient (BCG) attribute 1,2,3 taxa to BCG attribute 4,5,6 taxa < 0.75 with the ratio of BCG attribute 1,2,3 individuals to BCG attribute 4,5,6 individuals < 0.75?** This screening question evaluates the balance of pollution tolerant organisms with more sensitive organisms in terms of taxonomic richness and organismal abundance. By using the BCG attributes to measure pollution tolerance, this screening question serves as a check against the IBI metrics which account for pollution sensitivity based only on PTVs.

4. **Does the sub-sample show signatures of acidification year-round?** The primary acidification signatures in a sub-sample include low mayfly abundance and low mayfly diversity (i.e., scarce mayfly individuals and few mayfly taxa), especially when combined with high abundance of Amphinemura and/or Leuctra stoneflies, occasionally combined with high abundance of Simuliidae and/or Chironomidae individuals. A sub-sample with < 3 mayfly taxa, < 5% mayfly individuals, and > 25% Leuctra and/or Amphinemura stoneflies indicates likely acidification impacts. Acidification effects on benthic macroinvertebrate communities are often most pronounced in small streams with low buffering capacity during the spring months when snowpacks melt and vernal rains are frequent. While it can be difficult to determine if low pH conditions in a stream are natural or more attributable to anthropogenic acidification, sampling of water chemistry and/or fish communities (see Appendix F of PADEP 2009b) in addition to benthic macroinvertebrate communities can help inform assessment of acidic in-stream conditions. With this protocol, PADEP will only impair sites that show persistent acidification signatures year-round. In other words, if a sample has no mayflies and is dominated by Leuctra and Amphinemura in the spring, but a November sample from the same site contains three or more mayfly taxa or over five percent mayfly individuals, the aquatic life use will not be considered impaired because the stream exhibits the ability to recover biological integrity in the fall and winter months. If a spring sample shows acidification signatures, a late fall or early winter sample must be collected before making an aquatic life use assessment decision.

If the answer to any of these four screening questions is yes for a sample collected between November and May with an IBI score ≥ 50, then the sample is considered impaired without compelling reasons otherwise. If the answer to all of these questions is no for a sample collected between November and May with an IBI score ≥ 50, then the aquatic life use represented by the sample can be considered attaining unless other information (e.g., water chemistry) indicates the aquatic life use may not be fully supported at that location.

For samples collected between June and September, the same logic applies as for samples collected between November and May, but the attainment/impairment

threshold is lowered to 43 instead of 50. For samples collected in the summer and early autumn time frame, the absence of mayflies – and in some instances stoneflies – in samples collected immediately after seasonal hatches may be relaxed in some cases. Because benthic diversity may be underrepresented in summer and early autumn samples PADEP encourages monitoring in the November to May timeframe if possible. Benthic macroinvertebrate sampling for determining aquatic life use support should only be conducted from June to early October if sampling during other seasons is not possible due to hazardous conditions such as high, fast stream flow.

Limestone Influence

As discussed in the introduction, PADEP deploys a different sampling methodology and assessment protocol for limestone spring streams whose flow is mostly or entirely derived from groundwater in areas with substantial primary calcareous geologies (PADEP 2009a) than for freestone streams. The sampling methodology and assessment protocol for these limestone spring streams incorporate the understanding that streams in areas receiving a substantial amount of flow from groundwater attributable to karst geologies often naturally have less diverse benthic macroinvertebrate communities than streams draining freestone geologies. This lower benthic macroinvertebrate community diversity in limestone spring streams is attributable in large part to less variable flow and thermal characteristics of such systems when compared with freestone streams that often exhibit flashier flows and a wider range of temperatures.

Some streams in Pennsylvania drain basins underlain partially by freestone geologies and partially by calcareous geologies. Such streams are often encountered in central regions of the state – especially in upper portions of the Juniata River basin – where they drain sandstone and/or quartzite upland ridges, fairly steep shale slopes, and lower gradient calcareous valley floors. The calcareous valley geologies in these basins contributes to relatively high alkalinities and relatively high and consistent base flows in streams – characteristics of limestone spring streams – when compared with streams draining basins with no calcareous geologies. However, the upland sandstone, quartzite, and shale areas of these basins often contribute substantial surface runoff, which leads to surges in flow during rainfall and snowmelt events and dilution of alkalinity derived from the calcareous valleys. These streams – often referred to as “limestone-influenced” – exhibit some characteristics of limestone spring streams and some characteristics of freestone streams.

We often see substantial agriculture in the fertile valleys of these limestone-influenced streams, which makes it difficult to definitively establish reference conditions specific to these unique streams. However, there is evidence that the benthic macroinvertebrate communities in limestone-influenced streams are naturally less diverse than in freestone streams of similar size and with similar land uses. This lower diversity of benthic macroinvertebrate communities in limestone-influenced streams likely reflects the less variable flow and thermal patterns in these streams caused by the stabilizing influence of the substantial groundwater flowing into the streams through the calcareous

valley geologies. Commonly, the benthic macroinvertebrate communities in limestone-influenced streams exhibit relatively low stonefly diversity and abundance when compared with streams of similar size and condition that drain freestone geologies.

In light of these considerations, use attainment benchmarks may be justifiably relaxed for samples from limestone-influenced streams. The June to September IBI benchmark of 43 for freestone streams can be applied to limestone-influenced streams year-round, but the four screening question should still be applied as outlined above to samples from limestone-influenced streams to make ALU assessment decisions.

Antidegradation, Special Protection Considerations

The assessment decision process is somewhat different for streams with special protection uses of high-quality (HQ) or exceptional value (EV) waters. PADEP will protect special protection streams based on a baseline IBI score determined by previous surveys. Subsequent samples from HQ and EV streams will be compared to the baseline IBI score for a given site using the IBI temporal precision estimates (see table below). For example, if Riverkill Creek is designated HQ and a previous sample from a given site on Riverkill Creek using the protocol described above results in a mid-April IBI score of 78.0, this IBI score of 78.0 would be the baseline IBI score for that site. Future samples from that site collected November to May that score more than 10.0 IBI points below 78.0, would be considered impaired. Since PADEP's sampling season for special protection surveys is November to May, we need not be concerned about how June to October samples compare to the baseline IBI – PADEP will only make assessment decisions for HQ and EV streams based on samples collected November to May. The temporal precision estimate of 10.0 points is used because it approximates the October to May temporal precision estimate calculated in the table below. PADEP will apply the more restrictive March to May and October to February temporal precision estimates – about 9.0 and 8.0 IBI points, respectively – to special protection use assessments if the situation is appropriate (e.g., if the baseline IBI was established in April, future March to May samples that score more than 9.0 points lower than the baseline will be considered impaired). Furthermore, any sample from an HQ or EV stream that scores less than 63.0 on the IBI will be considered impaired without compelling reasons otherwise (e.g., a stream was designated HQ or EV for a reason other than assessment of the benthic macroinvertebrate community).

Metric		small-stream						large-stream					
		November to May 384 samples from 137 sites			June to September 26 samples from 12 sites			November to May 78 samples from 26 sites			June to September 26 samples from 7 sites		
		ANOVA MSE	90% CI (1 sample)	CV	ANOVA MSE	90% CI (1 sample)	CV	ANOVA MSE	90% CI (1 sample)	CV	ANOVA MSE	90% CI (1 sample)	CV
IBI score		48.9	**8.96**	**8.8%**	95.7	**12.54**	**19.6%**	69.0	**10.65**	**10.3%**	18.5	**5.51**	**4.8%**
Total Taxa Richness	*s*	115.0	13.75	10.9%	101.0	12.88	13.3%	128.0	14.50	12.5%	103.0	13.01	10.0%
	r	16.6	5.22	13.2%	16.1	5.14	14.8%	15.5	5.05	13.2%	12.1	4.46	11.3%
EPT Taxa Richness (PTV 0-4 only)	*s*	138.0	15.06	18.5%	89.5	12.13	23.8%	185.0	17.44	17.3%	78.8	11.38	10.7%
	r	6.3	3.21	19.7%	4.8	2.81	24.7%	7.9	3.59	20.8%	2.0	1.82	10.7%
Beck's Index (version 3)	*s*	127.0	14.45	22.8%	94.4	12.46	36.9%	132.0	14.73	14.2%	142.0	15.28	24.6%
	r	21.9	6.00	23.7%	17.9	5.42	37.5%	16.0	5.13	19.7%	10.4	4.13	26.4%
Hilsenhoff Biotic Index	*s*	53.1	9.34	7.3%	222.0	19.10	22.6%	71.3	10.83	8.3%	18.5	5.51	4.5%
	r	0.4	0.79	15.6%	1.5	1.57	21.2%	0.4	0.81	15.4%	0.1	0.38	6.1%
Shannon Diversity	*s*	96.1	12.57	10.1%	131.0	14.67	14.1%	120.0	14.04	10.5%	33.5	7.42	5.3%
	r	0.1	0.38	10.7%	0.1	0.45	14.4%	0.1	0.42	10.8%	0.0	0.24	5.7%
% Sensitive Individuals (PTV 0-3 only)	*s*	215.0	18.80	23.6%	361.0	24.36	65.7%	337.0	23.53	27.7%	133.0	14.78	16.5%
	r	157.0	16.06	23.8%	258.0	20.59	65.7%	197.0	23.53	30.2%	59.1	9.86	16.5%

Applications and Exceptions

If a sample results in fewer than 160 total organisms in the entire sample, the IBI and assessment procedures may not apply exactly as outlined above. The IBI and associated benchmarks are calibrated for use with sub-samples containing 160 to 240 organisms, so applications of the IBI to samples containing less – or more – than the target number of organisms, cannot necessarily be assessed using the procedures and benchmarks outlined above. Low abundance of benthic organisms often indicates toxic pollution or severe habitat alterations, which must be considered in making holistic stream assessments.

The use assessment decision processes set forth above are intended as general guidelines, not as hard-and-fast rules. The procedures and guidelines discussed above will provide tenable assessments – as required by federal and state law – of benthic macroinvertebrate community conditions for the vast majority of samples collected from wadeable, freestone, riffle-run streams in Pennsylvania. However, as noted by Hughes (1995), there will be exceptional circumstances – such as those outlined in the Pennsylvania Code (2011: Title 25, Section 93.4.(b) relating to less restrictive uses) – when the above assessment procedures do not apply (e.g., there are no obvious sources of impairment and natural factors such as habitat availability or water chemistry limit biotic potential). In some situations a biologist's local knowledge of conditions may warrant a decision not arrived at using these guidelines. Although the large-steam IBI appears to work fairly well when applied to samples from large rivers (i.e., sites draining over 1,000 square miles), discretion must be used when applying this IBI to samples from such large rivers (these methods do not apply if a stream/river is not wadeable in over 90% or more of its channel area under base flow conditions for the river segment to be sampled or other situations not consistent with riffle and run dominated habitat). The relatively small dataset of samples from such large rivers used in the IBI development limits analysis of variability (i.e., estimates of spatial and temporal precision) in metric and IBI performance with samples from such large rivers.

In other situations, like when samples are heavily dominated by Prosimulium larvae – as discussed above – often times this will unduly lower metric and IBI scores, confounding the assessment decision procedures outlined above. In such situations, the investigating biologist may have to re-sample the site after the seasonal Prosimulium larval boom, or the biologist may have to rely on a more qualitative analysis of metric scores, sample composition, and site conditions to arrive at an assessment decision. In any instance, evaluating stream samples requires mindfulness of particular conditions, and is not always a definite, exact exercise. A certain section of stream may represent a transition between pool-glide, low-relief, marshy, glaciated uplands where the substrate is mostly fine-grained sand and higher-gradient lower reaches filled with cobble-strewn riffles and runs. Some years see cooler, wetter springs than other years.

Nevertheless, for the vast majority of cases involving benthic macroinvertebrate samples from wadeable, freestone (and limestone-influenced), riffle-run streams in Pennsylvania using the protocols described above, the assessment procedures described in this report will lead to tenable ALU assessment decisions.

REFERENCES

Barbour, M.T., J.B. Stribling, and J.R. Karr. 1995. Multimetric approach for establishing biocriteria and measuring biological condition. Chapter 6 in *Biological assessment and criteria: tools for water resource planning and decision making*, W.S. Davis and T.P. Simon, eds. (pp. 63 – 77). CRC Press, Boca Raton.

Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish, second edition. EPA 841-B-99-002. United States Environmental Protection Agency; Office of Water. Washington, D.C.

Beck, W.H., Jr. 1955. Suggested method for reporting biotic data. Sewage and Industrial Waste 27(10): 1193-1197.

Davis, W.S. and T.P. Simon. 1995. Introduction to *Biological assessment and criteria: tools for water resource planning and decision making*, W.S. Davis and T.P. Simon, eds. (pp. 3 – 6). CRC Press, Boca Raton.

Davies, S.P. and S.K. Jackson. 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. Ecological Applications 16(4):1251-1266.

Hawkins, C.P. 2006. Quantifying biological integrity by taxonomic completeness: its utility in regional and global assessments. Ecological Applications 16(4): 1277-1294.

Hilsenhoff, W.L. 1977. Use of arthropods to evaluate water quality of streams. Technical Bulletin Number 100. Wisconsin Department of Natural Resources. 15 pp. Madison, Wisconsin.

Hilsenhoff, W.L. 1987. An improved biotic index of organic stream pollution. The Great Lakes Entomologist 20(1): 31-39.

Hilsenhoff, W.L. 1988. Rapid field assessment of organic pollution with a family-level biotic index. Journal of the North American Benthological Society 7(1): 65-68.

Hughes, R.M. 1995. Defining acceptable biological status by comparing with reference conditions. Chapter 4 in *Biological assessment and criteria: tools for water resource planning and decision making*, W.S. Davis and T.P. Simon, eds. (pp. 31 – 47). CRC Press, Boca Raton.

Klemm, D.J, P.A. Lewis, F. Fulk, and J.M. Lazorchak. 1990. *Macroinvertebrate field and laboratory methods for evaluating the biological integrity of surface waters*. Environmental Monitoring Systems Laboratory, United States Environmental Protection Agency. Cincinnati, Ohio. EPA-600-4-90-030.

Lenat, D.R. and D.L. Penrose. 1996. History of the EPT taxa richness metric. Bulletin of the North American Benthological Society 13(2).

Pennsylvania Department of Environmental Protection. 2009. Survey Protocol – Instream Comprehensive Evaluation Surveys. 391-3200-001. Harrisburg, PA.

Pennsylvania Department of Environmental Protection. 2003. Standardized Biological Field Collection and Laboratory Methods. Harrisburg, PA.

Shannon, C.E. 1948. A mathematical theory of communication. Bell System Technical Journal 27: 379-423 and 623-656.