Water Quality Assessment/Policy 1-11 Updates
Bioassessment Alternatives Proposed by Ecology
January 2017

Ecology is proposing the following 3 Alternatives for purposes of public dialogue on the use of bioassessment data for listing in the Water Quality Assessment:

- **Draft Alternative #1:** Use two bioassessment thresholds to delineate into degraded, non-degraded, and inconclusive (similar to current methodology).

- **Draft Alternative #2**: Use a single bioassessment threshold to delineate degraded from non-degraded.

- **Draft Alternative #3:** Don't use bioassessment scores as thresholds, but instead treat individual community metrics and/or characteristics as independent lines of evidence in making impairment determinations, without relying on overall B-IBI or RIVPACs scores.

Features Common to All 3 Alternatives:

Prior to assessing bioassessment scores, the number of organisms in each sample would be evaluated. The goal is to evaluate a benthic macroinvertebrate assemblage based on samples with at least 500 individual organisms. However, having fewer than 500 organisms in a sample does not automatically indicate compromised data quality, such as inadequate completeness or representativeness. Samples with less than 500 organisms can indicate suboptimal sampling effort, unproductive habitats, severe water quality problems, and/or severe habitat degradation. In general, samples sizes below approximately 300 organisms provide cause for evaluating the potential reasons that the sample does not attain the desired organism count. Samples containing less than 300 organisms would be further evaluated in order to ascertain the probable reason for the suboptimal count. A sample with less than 300 organisms may be rejected if it is attributed to sub-optimal sampling effort. Whether or not a sample is rejected based on the organism count will depend on the known or probable condition of the waterbody being evaluated. Bioassessment scores from sites with depressed taxa abundances would be used in the impairment evaluation, as they are indicative of potential degradation. If a site has a history of not achieving samples with 300 or more organisms, this likely would not indicate sub-optimal sampling effort, but rather depressed taxa abundance due to one or more of the factors listed above. Ecology has observed samples with fewer than 300 organisms that resulted in relatively high B-IBI scores. Conversely, we have also observed samples with fewer than 300 organisms that resulted in low B-IBI scores.

All alternatives would employ the same basic scheme that was used in the current EPA approved assessment to categorize the aquatic life use of waterbodies as impaired or non-impaired based on bioassessment data. Benthic assemblages vary seasonally and inter-annually due to complex interactions among site to landscape scale chemical, physical, and biological factors. Recent analyses have indicated that B-IBI scores may fluctuate inter-annually on average by 10%. To account for such uncertainty and increase confidence in the final category determination, data from two or more years will continue to be required to determine whether

or not impairment of the aquatic life use of an assessment unit is indicated. For listings that existed in a prior assessment, the listing category would generally only be changed if new data justifies a change in category.

- Category 1 (evidence indicates non-impairment of the aquatic life use): the two most recent years of data indicate a non-degraded benthic community.

- Category 5 (evidence indicates impairment of the aquatic life use): two of the five most recent years with data indicate a degraded benthic community.

- Category 2 (water of concern): the available data in the 10 year assessment window does not conclusively indicate a degraded or non-degraded benthic community (i.e. the data does not qualify for Category 1 or 5). Data from a single year indicating a potentially degraded benthic community would result in a Category 2 listing.

- Category 3 (insufficient information): only one year of bioassessment data is available and the data indicates a non-degraded community.

Bioassessment results that indicate a non-degraded community in the two most recent years with data would be required to place a listing in Category 1. This would provide certainty that a non-degraded benthic community is the most recent condition of the waterbody given that B-IBI scores may fluctuate form year to year by about 10% on average.

To qualify for Category 5, results indicating degradation from two or more years in the five most recent years with data would be required. Within the 10 year water quality assessment data window, up to five years of bioassessment data would be considered for Category 5 determinations. However, a minimum of two years of data may be sufficient to place a listing in category 5. Note that the difference between considering *the most recent five years* and *the five most recent years with data* is that in the latter case data from non-consecutive years may be considered. This is an important distinction because bioassessment data are often not available for multiple consecutive years. The Category 5 requirements that the indication of degradation is reflected in the most recent data available and that it has persisted across at least two years should help to avoid categorizing waterbodies as impaired in cases where biological integrity started off with poor scores early in the 10 year data window, but improved later in the 10 year data window due to watershed protection efforts, for example.

Listings would qualify as a water of concern (Category 2) if they do not qualify for Categories 1, 5, or 3. Category 2 would result when an assessment unit has data from one or more years in the 10 year data window that indicates degradation or potential degradation, but the dataset overall does not conclusively indicate impairment or non-impairment. Under these proposals it would be possible to move from a Category 5 to a Category 2 or 1 based on data assessed within the ten year window.

Significant Changes Proposed by Ecology from the Current Assessment Methodology (would apply to Alternatives A & B)

One significant change from the current bioassessment method that would be included in Alternatives A and B would be that instead of using the lowest bioassessment score from a given year to represent the year, scores based on representative samples would be averaged together if more than one score is available. If RIVPACS, B-IBI, and/or MMI scores are available for a given year, they would all be converted into scores under a single model based on the raw assemblage data.

A second significant change would be that the B-IBI model would be applied for each EPA Level 3 ecoregion across the entire state, instead of only for western Washington as was done in the last assessment. The B-IBI scoring scale would be recalibrated from the 10 - 50 scale to a 0 - 100 scale and the B-IBI score thresholds would vary by ecoregion according to the distribution of reference site scores in a given ecoregion. Table I below approximates the 5th and 25th percentiles for reference sites within the various ecoregions of Washington using data through 2015; these percentile are shown for both the current and new scoring scales.

Table 1: *5th and 25th percentiles of B-IBI scores for Dept. of Ecology reference sites in the various EPA level III ecoregions in Washington. Data are through 2015.*

			B-IBI (10-50)		B-IBI (0-100)	
Level III Ecoregion		**# visits**	**5th**	**25th**	**5th**	**25th**
Western Washington	**North Cascades**	36	30.00	36.00	57.02	74.75
	Cascades	31	32.00	40.00	56.85	82.75
	Coast Range	38	32.00	36.00	59.77	71.05
	Puget Lowland	28	26.10	36.00	49.84	77.62
	Willamette Valley[1]	0	NA	NA	NA	NA
Eastern Washington	**Eastern Cascades Slopes & Foothills**	28	25.40	34.00	43.95	67.10
	Northern Rockies	35	30.80	36.00	56.01	68.45
	Blue Mountains	17	31.20	40.00	56.32	79.90
	Columbia Plateau	31	18.00	27.00	26.80	52.00
All Ecoregions Combined		244	26.30	34.00	44.86	69.55

[1]The very small portion of the Willamette Valley Ecoregion in Washington would be combined with the Puget Lowland ecoregion for assessment purposes.

A fourth significant change would include a category modifier intended to address the stakeholder concern that some sites with historically high bioassessment scores may become steadily degraded before they are considered to have impaired biological integrity. To remedy this issue, a "safety net" would be established for waterbodies with historically high B-IBI scores. For example, a drop in scores of 15% relative to historic scores could cause a listing to go into Category 5, assuming that the waterbody doesn't already qualify for Category 5. In evaluating a 15% reduction in scores, the baseline would be the average of the two earliest historic scores and the comparison would be against the average of the two most recent scores. For listings initially based on RIVPACS scores, the B-IBI model would be applied to the datasets originally used to derive the RIVPACs scores.

<u>Alternative A:</u> Use two bioassessment thresholds to delineate into degraded, non-degraded, and inconclusive (similar to the current methodology)

Alternative A would retain the basic structure of the current bioassessment methodology. There would continue to be three condition classes- non-degraded, inconclusive, and degraded. The purpose of having three condition classes would be to address uncertainty in the biological integrity condition where there is considerable overlap in the distribution of the assessment site scores with the lower distribution of the reference scores.

Degraded Community

The threshold for a score indicating a degraded benthic community would be set at the 5th percentile of reference site scores (sites considered to be least-impacted by human influence). This means that 5% of reference site scores would be classified as degraded under this approach. In using reference sites to evaluate the biological integrity of non-reference sites it is considered impractical to set the impairment threshold below the 5th percentile. From an error analysis perspective, we are stating that we designate a site score as indicating degradation only when we are highly confident that we would not incorrectly conclude that an assemblage is degraded. The 5th percentile approximates what was used in the last water quality assessment as the threshold for degraded sites, which was a score below 28 on the 10 - 50 B-IBI scale. For B-IBI data through 2015, approximately 31% of random site scores from western Washington sites were below the 5th percentile of reference site scores (a score of 54) on the 0 -100 scale. It is interesting to note that no B-IBI reference site scores are below 40 on the 0-100 scale, whereas 21% of scores from random sites fall below a score of 40.

In Montana and Kentucky, the 10th percentile of the reference site scores has been used as the threshold for indicating degradation using multi-metric indices. In New Mexico, the 25th percentile of the reference site scores has been used as the threshold. In Oregon, the threshold for indicating impairment using a RIVPACS model is the 10th percentile of reference site scores; scores between the 10th and 25th percentile are considered to indicate a moderate level of disturbance and result in a 303(d) listing if further data and information implicate anthropogenic disturbance.

Reference sites with scores below the 5th percentile may be:

1) Indicative of a site visit in which sampling effort was insufficient to obtain a representative sample of the site;
2) Reflective of a recent cataclysmic disturbance such as a debris flow;
3) Strongly influenced by naturally occurring conditions that constrain the diversity and abundance of the benthic community; and/or
4) Misclassified as being least-impacted by human activities.

Conditions that may occur **naturally** in streams that could result in low bioassessment scores include:

- Recurrent high water temperatures (e.g. >23°C) or temperatures that are perpetually very cold (e.g. <8°C);
- A substrate dominated by fine sediment;
- A substrate dominated by bedrock;
- Flow impermanence (i.e. ephemeral or intermittent flow);
- Elevated toxin concentrations (e.g. naturally occurring metals); or
- High acidity waters (i.e. pH below 5.0); v
- Very low productivity (i.e. water with low ionic strength and/or low nutrient levels).

Inconclusive Determination

B-IBI scores between the 5thand 25th percentiles of reference sites scores would be classified as inconclusive. To put this into perspective, it means that 20% of reference site scores would be classified as reflecting inconclusive biological integrity. Designating a range of scores as inconclusive rather than degraded or non-degraded is intended to incorporate considerations of uncertainty in the comparison between reference and non-reference sites. Due to substantial overlap in the distribution of reference and non-reference scores, there is a range of scores in which the biological integrity of random sites is not clearly distinguishable from reference sites. In other words, we would be stating that based on the B-IBI scores alone, we are not confident enough to conclude that non-reference sites scores between the 5thand 25th percentiles of reference site scores have degraded biological integrity. This is because there are benthic assemblages at some least impacted reference sites that are similar to assemblages at random sites that are perhaps moderately degraded relative to their historic condition (possibly due to the some of the potentially natural reference site conditions noted previously). For B-IBI data through 2015, approximately 26% of random site scores from western Washington sites were between the 5th and the 25th percentile of reference site scores on the 0 - 100 scale.

Non-degraded Community

B-IBI scores above the 25th percentile of reference sites scores would be classified as non-degraded. This means that 75% of reference site scores would qualify as indicating a non-degraded community. We can say with a high degree of confidence that scores from random sites that are above the 25th percentile of reference site scores have biological integrity that is not significantly different than sites with apparently minimal human influence. For B-IBI data through 2015, approximately 43% of random site scores from western Washington sites were above the 25th percentile of reference site scores on the 0 - 100 scale.

Alternative A would include an important category modifier that is absent in the current method. This modifier is intended to address the stakeholder concerns that: 1) having separate thresholds to designate degraded vs. non-degraded scores may allow some sites to become steadily degraded before they are considered to have impaired biological integrity; and 2) the natural condition of some waters may lead to low B-IBI scores, but such scores may not be reflective of a pollution problem.

Proposed Use of Ancillary Data

For any listings that qualify for Category 2 or 5, ancillary community metrics (such as the fine sediment index, metals tolerance index, biotic index, temperature tolerance index, etc.) would be evaluated against reference sites in order to determine if a category modification is appropriate. In other words, multiple low scores would trigger further analysis of the community. For purposes of comparison, an ancillary metric score would be considered different if it is below the 10th percentile of the distribution of the reference site scores for that metric. Using the community data that resulted in the lowest B-IBI scores:

- If any ancillary metric for a candidate Category 5 listing is different than those of reference sites, then the listing would be assigned to category 5.

- If all ancillary metrics for a candidate Category 5 listing are <u>not</u> different than reference sites values, then place listing in Category 2.

- If any ancillary metric for a candidate category 2 listing is different than reference sites values, then place listing in Category 5.

- If all ancillary metrics for a candidate Category 2 listing are <u>not</u> different than reference sites values, then place listing in category 1.

This modifier does not remedy the problem where sites have low scores due to naturally high fine sediment, metals, temperature, etc. However, it does help to increase the confidence in the final category determination. Only further investigation which characterizes the natural condition of a waterbody can distinguish between waters having naturally high pollutant levels and those waters where humans have caused elevated pollutant levels.

<u>Alternative B:</u> Use a single bioassessment threshold to delineate degraded from non-degraded sites, instead of having range of score for potentially degraded sites.

Alternative B would modify the basic structure of the current bioassessment methodology to address the concerns that: 1) having separate thresholds for degraded vs. non-degraded sites may allow some waters to become steadily degraded before they are identified as impaired; and 2) impaired sites that have scores ≤ 27 have to subsequently achieve scores of 38 or above before they can be de-listed. There would be two condition classes, non-degraded and degraded, instead of three classes. Uncertainty in bioassessment scores due to overlap in the distribution of reference sites and assessment sites having sub-optimal scores would be addressed in a manner similar to Alternative A.

The single threshold for delineating the two condition classes (Degraded and Non-degraded) would be the **10th percentile** of the reference site score distribution. The 10th percentile of reference site scores has been used as a threshold for indicating degradation in Montana, Kentucky, and Oregon. Category 2 determinations would be made using ancillary data, described below.

Proposed Use of Ancillary Data

Alternative B would include an important category modifiers that is absent in the current method. These modifiers seek to address the stakeholder concerns that: 1) that waters may become steadily degraded before they are considered to have impaired biological integrity; and 2) the natural condition of some waters may lead to low B-IBI scores.

For any listing that qualifies for Category 5, ancillary community metrics (such as the fine sediment index, metals tolerance index, biotic index, temperature tolerance index, etc.) would be evaluated against reference sites in order to determine if a category modification is appropriate. For purposes of comparison, an ancillary metric score would be considered different if it is below the 10th percentile of the distribution of the reference site scores for that metric. Using the community data that resulted in the lowest B-IBI scores:

- If all ancillary metrics are <u>not</u> different than reference sites values, then place listing in Category 2.

This category modifier does not remedy the problem where sites have low scores due to naturally high fine sediment, metals, temperature, etc. However, it does help to increase the confidence in the final category determination. Only further investigation which characterizes the natural condition of a waterbody can distinguish between waters having naturally high pollutant levels and those waters where humans have caused elevated pollutant levels.

<u>Alternative C</u>: Don't use bioassessment scores as thresholds, but instead treat individual community metrics and/or characteristics as independent lines of evidence in making impairment determinations.

This alternative would address the concern that the cumulative B-IBI score may not actually indicate impairment, for example, where sites are outside of the experience of the model. An Ecology staff biologist would compare the community observed at a site to the community at reference sites deemed to be sufficiently comparable in terms of natural chemical, physical, and biological attributes. This would require a biologist to review the community data for each monitoring site relative to historical data for the location or relative to reference conditions in order to develop conclusions about whether or not the community at that site is healthy (Category 1). If any community aspect were deemed to not be in the appropriate range for the site of interest then the listing would be assigned to Category 5. There would be no pre-established metrics that would need to be evaluated consistently across all sites relative to reference sites. For example, if the percentage of the community that is tolerant to pollution appears to be significantly elevated above reference site values, or the number of mayflies, caddisflies, and stoneflies appears to be depressed, then the site could be listed based on the best professional judgement of the biologist that the site is impaired.

This alternative would provide much more flexibility in making listing decisions, but would not result in consistency across sites. For example, one site that has a low number of mayfly, stonefly, and caddisfly (EPT) taxa may not be placed in Category 5 if it is determined to be similar enough to reference sites, while another site with the same number of EPT taxa may be placed in Category 5 if it were determined by a staff biologist that the EPT taxa number is too low and thus not indicative of a healthy system. Since knowledge would be incomplete for every community aspect and individual sites, the biological expertise of the assessor would be crucial to correctly determining whether or not the observed community at a site is similar enough to the community that would be expected for that site under non-impaired conditions. Among the three alternatives, this method would require more detailed descriptions of the analysis and methodology needed in order to ensure some consistency across ecoregions. Because this alternative would not use a modeled approach with reference sites as a baseline, it could result in introducing Type 1 (incorrectly determining a waterbody is impaired) and Type 2 errors (incorrectly determining a waterbody is not impaired). It is also unclear when or if Category 2 determinations would be made under this alternative.