

Integrated Report Improvements

Category 3B and Overwhelming Evidence

Date: May 11, 2018

Introduction and background

Section 303(d) of the CWA requires states to prepare lists of “surface waters that do not meet applicable water quality standards”, referred to as the impaired waters list, and to establish Total Maximum Daily Loads (TMDLs) for pollutants causing the impairment of these waters on a prioritized schedule. The 303(d) list is the state’s opportunity to utilize scientifically and statistically valid information to assess waterbodies and prioritize the waterbodies and pollutants for TMDL development.

EPA recommends using five reporting categories to classify water quality status. The categories represent varying levels of water quality standards attainment and beneficial use support, ranging from Category 1, where all designated uses for a water body are supported, to Category 5, where a water body is impaired and a TMDL is required to return the water to a condition where the water quality standards are met. Historically, in the absence of robust data sets and in an effort to be protective of beneficial uses, DEQ has used a conservative, low threshold for placing waterbodies in Category 5 on the state’s 303(d) list.

DEQ’s Assessment Methodology contains documentation required by federal regulations to support listing and delisting determinations that is submitted to EPA along with DEQ’s final updates to Oregon’s 303(d) list. The Assessment Methodology describes how DEQ compares water quality data to Oregon’s water quality standards, which are adopted in rule in OAR 340-041. The Methodology does not constitute a “rule”, but describes how DEQ interprets Oregon rules (i.e., water quality standards) in order to evaluate water quality data for the purpose of assessing state waters and identifying water quality impairments.

The primary factor for determining an impairment listing is the comparison of available, and many times limited, data on water quality conditions with the applicable water quality criterion. When sampling frequency is low, and sample sizes are small, criteria excursions may be interpreted as the water body exceeding water quality standards criteria and not supporting its beneficial uses. On the other hand, a small number of attaining samples is also used as the basis to conclude the water body is attaining standards and supporting designated beneficial uses when this small number might not be enough data to reliably capture critical conditions and represent variability in the waterbody. Interpreting small amounts of data has high uncertainty and poorly minimizes errors in making either attainment or impairment decisions.

In the case of limited data sets, the concept of “overwhelming evidence” has routinely been used by other states for making decisions that waterbodies are impaired using more information than just the number of samples available. Intermediate listing categories, such as Category 3B, that signify uncertain status between attainment and impairment can also be used to identify waters that cannot be reliably considered either impaired or attaining given the limited data at hand.

EPA’s 2002 “Consolidated Assessment and Listing Methodology” states:



State of Oregon
Department of
Environmental
Quality

Water Quality Standards and Assessments

700 NE Multnomah St.
Suite 600
Portland, OR 97232
Phone: 503-229-5696
800-452-4011
Fax: 503-229-5850
Contact:
Becky Anthony

www.oregon.gov/DEQ

DEQ is a leader in restoring, maintaining and enhancing the quality of Oregon’s air, land and water.

An assessment methodology should take into account the balance between desired data requirements and the practical realities affecting the availability of information and the strength of the available evidence. Generally, decisions should be based on very small sample sizes only when there is overwhelming evidence for impairment. EPA does not recommend making decisions based on small sample sizes of water column chemistry for attainment.¹

Overwhelming evidence leading to impairment conclusion

DEQ is often faced with the need to make decisions about impairment based on limited data. One approach used by other states to address listing decisions when sample sizes are minimal but there is additional information that impairment is likely, is the concept of “overwhelming evidence” (Table 1).

Overwhelming evidence uses multiple lines of evidence based on specific rationale to conclude waterbodies are impaired. When sample sizes do not meet minimum requirements to assign a Category 5 status, additional evidence can be used to indicate that the applicable water quality standard is not being met. Overwhelming evidence includes other credible and compelling information indicating the waterbody is in fact impaired.

Several states outline the types of evidence they consider when concluding that there is overwhelming evidence of impairment (Table 1). Evidence may be other characteristics of the data at hand, such as occurrence of extremely high values, or other types of observation about the conditions in the waterbody. These additional observations do not necessarily need to be associated with specific numeric water quality standards.

Table 1. Examples of the use of “overwhelming evidence” in other states

State	Overwhelming Evidence of Impairment Language
Idaho (R10)	<ul style="list-style-type: none"> • A single valid data point \geq than 2 times chronic criterion*
Alaska (R10)	<ul style="list-style-type: none"> • <i>In progress</i> - Formalizing in methodology • Case-by-case basis • Multiple and frequent excursions are corroborated by nearby sampling stations • Confirmed pollutant sources • Biological impacts verify chemical impacts
Montana (R8)	<ul style="list-style-type: none"> • \geq 1 sample exceeds twice the acute aquatic life water quality standards (WQS) • \geq 2 exceedances of aquatic life WQS within an existing sample size of n = 3 to 7 • Other approaches vary by parameter
Colorado (R8)	<ul style="list-style-type: none"> • Sufficient and credible data that clearly demonstrate a waterbody's designated uses are impaired and used when minimum data requirements are not met • Exceedance of numeric water quality standards by more than 50 percent in magnitude.

¹ EPA, 2002. Consolidated Assessment and Listing Methodology, First Edition, July 2002. Prepared By: U.S. Environmental Protection Agency Office of Wetlands, Oceans, and Watersheds.

State	Overwhelming Evidence of Impairment Language
Iowa (R7)	<ul style="list-style-type: none"> • Presence of recurring, man-made circumstances resulting in acutely toxic conditions • Man-made alterations of hydrology, flow or habitat • Chronic de-watering • Presence of exotic species • Summer median Trophic State Index (TSI) based on < 3 years that is more than 5 TSI points greater than criteria value • <i>E. coli</i> geometric mean of at least 5 samples exceed Iowa's geometric mean criteria even if remainder of samples do not exceed criteria
New Hampshire (R1)	<ul style="list-style-type: none"> • "Magnitude of Exceedance Criteria" (MAGEXC) established for many parameters. • ≥ 2 samples exceed the MAGEXC, waters are assessed as impaired, regardless of the total number of samples taken.
New Jersey (R2)	<ul style="list-style-type: none"> • Datasets smaller than the target sample size • Include data from nearby chemical or biological sampling station(s) upstream or downstream verify similar conditions; • Data collected prior to the last 5 years at the sampling station support the assessment decision; • Data from confirmed pollutant sources verify impacts; • Trends corroborate with current water quality; • Hydrologic conditions signify water quality impacts; • Biological conditions concur with water quality data; • Natural conditions validate the assessment decision
Minnesota (R5)	<ul style="list-style-type: none"> • "<i>Weight of evidence</i>"; quality and quantity of all available data; • Magnitude, duration and frequency of exceedances; • Timing of exceedances; naturally occurring conditions that affect pollutant concentrations and toxicity; • Weather and flow conditions; • Known influences on water quality in the watershed; • Changes in the watershed that may have changed water quality • Use "Professional Judgment Team"

*Two times the criteria is often used because of the way acute criteria (CMCs) for toxic substances are derived. A final acute value (FAV), which is expected to be lethal to 50% of sensitive species, is divided by two to obtain the CMC (i.e., $CMC = FAV/2$) (Stephan and others, 1985). This acute criterion derivation is performed to reduce a lethal concentration to a concentration expected to kill few, if any, organisms. It follows that if a reliably measured concentration is greater than twice the CMC (i.e., is greater than or equal to the FAV), it is likely to be lethal to sensitive organisms used in criterion development. Therefore, the assessor may conclude the water body is unlikely "free from toxics in toxic amounts" if an excursion has been documented at twice the CMC.

Category 3B

A second tool that is employed for assessing small sample sizes, where some samples indicate possible impairment, is the use of Category 3B. Category 3B is a sub-category of Category 3 used by many states (e.g. Colorado, Utah, Virginia, Iowa, and Missouri) that identifies water bodies that are potentially impaired but additional information needs to be collected to confirm the finding of impairment.

DEQ is considering the use of Category 3B under the following conditions/ circumstances:

- Exceedances of criteria occur but data do not meet minimum sample size or confidence required to list;
- Conflicting evidence of attainment;
 - Use of total to dissolved translators;
 - Calculated criteria where default input parameters (e.g. hardness-based metals criteria, biotic ligand model) are used;
- Use of estimated data; values below the reporting limit (MRL)
- Incomplete records of continuous monitoring data; continuous data does not represent critical periods

Insufficient sample size

Typically, water bodies are placed in Category 3B: Insufficient Data – Potential Concern until more data can be collected to verify the impairment decision. In cases of insufficient sample size where there are no sample excursions, attainment decisions default to Category 3.

Category 3B is currently used in DEQ’s assessment conclusions for possible impairments, but only where insufficient data exists (Table 2). DEQ is exploring additional scenarios and formalizing the use of Category 3B to increase the confidence in Category 5 determinations and minimize the occurrences of listing waterbodies without sufficient evidence of impairment.

Table 2. Examples of Category 3 in DEQ’s 2012 Assessment Methodology

Parameter	Data Requirement	IR Category
Toxic Substances	1 sample > criterion	3B
pH	When $n \leq 5$; ≥ 2 samples exceed criterion	3B
Dissolved Oxygen (grab)	When $n \leq 10$; < 10% samples exceed criterion	3
Dissolved Oxygen (continuous)	Insufficient data to calculate 30- and 7-day averages	3
Temperature	Insufficient data to calculate 7-day average max temperature	3

Conflicting evidence

Category 3B may also be applicable where there sufficient data exists, but conflicting attainment interpretations exist. These situations arise when there is uncertainty or a mismatch in the information available to assess attainment of water quality standards. DEQ will prioritize the use of measured input data and dissolved data where those data meet minimum data requirements. If minimum data requirements are not met, then DEQ will utilize all available data for the assessment

Situations where this might occur are:

- When samples measured as total recoverable exceed a dissolved criterion.
- In application of water chemistry-based criteria; when ALL samples with concurrent water chemistry date attain criteria , but enough samples without water chemistry data (and consequently based on default water chemistry values) are sufficient to list. This is a plausible scenario for
 - ammonia
 - hardness-based metals

- biotic ligand model for copper

To address the possibility of conflicting results among different types of data used to assess attainment with water quality standards, EPA recommends that states apply a policy of independent applicability.² The policy of independent application is based on the premise that any valid, representative dataset indicating an actual or projected water quality impairment should not be ignored when one is determining the appropriate action to be taken.³ The intent of this policy is to protect against dismissing valuable information when evaluating aquatic life use support, particularly in detecting impairment. When differences occur in attainment decisions, DEQ must determine whether the differences can be attributed to an artifact of the data set (e.g. analytical methods, sample representativeness, etc.).

The policy *does not* say that a single sample result showing impairment outweighs all of the other data demonstrating attainment. Therefore, circumstances may arise when conflicting attainment results should be investigated further before a final attainment or non-attainment decision is made. For example, states may obtain datasets from third parties of varying quality, which may influence the reliability of the assessment results. Placing the assessment unit in Category 3B would allow for more information to be collected, and greater certainty in the final decision to place the assessment unit in Category 5 or 2.

Through the process of data pooling in an Assessment Unit, there will be cases where two or more types of data do not indicate consistent attainment status. In such cases, further investigation may be warranted. A determination should be made whether differences in assessment results can be attributed to differences in the quality of the datasets, age of the datasets (i.e. historic versus recent) or other environmental or hydrologic factors. When the differences can be attributed to data quality issues (i.e. measured versus default data), the independent application policy allows for resolving the differences by giving the higher quality dataset more weight in the attainment decision. In cases where conflicting assessment results occur and data quality is not at issue (such as when dissolved metals data attain criteria while total metals data assessed against the dissolved criteria demonstrate exceedances), DEQ proposes to place the waterbody in Category 3B (requiring additional study or monitoring) to better understand and resolve the conflicting lines of evidence.

In past assessments, DEQ conservatively determined water quality impairment. Waterbodies were included in Category 3B only under limited circumstances. The method used in Oregon's 2012 303(d) Integrated Report to assess toxics interprets the frequency component of the standard very narrowly and requires only one valid sample result that does not meet the most stringent applicable criterion to place the segment in Category 3B.

Analysis – Possible Category 3B Scenarios

² EPA, 2002. Consolidated Assessment and Listing Methodology, First Edition, July 2002. Prepared By: U.S. Environmental Protection Agency Office of Wetlands, Oceans, and Watersheds.

³ EPA, 2005. Guidance for 2006 Assessment, Listing and Reporting Requirements Pursuant to Sections 303(d), 305(b) and 314 of the Clean Water Act. Prepared By: U.S. Environmental Protection Agency Office of Wetlands, Oceans, and Watersheds.

Scenario 1

Data for a given water body were collected by DEQ, USGS and a municipality. Data from USGS and DEQ were determined to be the highest quality (tier1), data from the municipality was placed in a tier 2 category. Tier 2 data can denote that the sample result is an estimated value because the concentration of the pollutant was detected, but at a concentration lower than the quantification limit. In cases where the numeric criterion is similar to the quantification limit of some analytical methods, the estimated concentrations can be counted as exceedances.

There were no exceedances of the criterion in the combined USGS/DEQ dataset, but several exceedances occurred in the municipality's dataset over the same period. In this example, more weight would be given to tier 1 data, therefore making the impairment decision inconclusive and a Category 3B determination appropriate. Monitoring resources could be dedicated to determine what the appropriate status should be. Rather than creating an incentive to remain in Category 3B indefinitely, DEQ could require that the water body be placed in Category 5 after a given number of Integrated Report cycles if no new data were collected.

Scenario 2

During the assessment process, DEQ obtains multiple datasets of varying quality. The form in which a pollutant was measured may not match the criteria, or water chemistry data needed to assess the pollutant may not always be available which may provide conflicting evidence of the assessment status of the waterbody. In some cases, numeric criteria values are dependent on water quality conditions. These include the criteria for many metals and ammonia. For hardness-based metals, assessment datasets may contain a mix of criteria calculated from hardness data (not sufficient to meet minimum sample size) and the use of defaults when hardness data were not collected. As default values are highly conservative, scenarios may arise where the data assessed against measured hardness are attaining while data assessed against the defaults are exceeding.

Similarly, most metals criteria are expressed in terms of the dissolved metal. Permitted effluent limits are expressed as total metals. In past assessments, total metals data collected by entities for permit compliance or historical total metals data were compared to dissolved metals criteria and placed in Category 5 if exceedances occurred. In the next assessment, total metals translators will be used for some pollutants, but if exceedances occur that are contradicted by dissolved metals data, the segment would be placed in Category 3B. When different assessment outcomes can be attributed to data issues, differences may be resolved by weighing the datasets which ultimately factors into the assessment decision.

Conclusion / Recommendation

Based on the examples presented above, the recommendation of the IR Improvement Team is to expand the use of Overwhelming Evidence and Category 3B under "Section D. Determining Impairment Status" in the revised Assessment Methodology. During the assessment process, DEQ will evaluate all factors such as magnitude of exceedance, critical time periods and additional lines of evidence when making impairment decision. Although DEQ has tried to anticipate all cases where overwhelming evidence or Category 3B may be used, this is not an exhaustive list. There will be cases that fall outside of the scenarios that have been laid out and DEQ will address them on a site-specific basis. Accumulation of assessment experience will continue to inform and contribute to future revisions to DEQ's assessment methodology.

DEQ would adopt the following guidelines for indicators of Overwhelming Evidence:

Extreme exceedance of criteria	<ul style="list-style-type: none"> • Samples exceed at 2x the acute magnitude
Corroboration with nearby Assessment Units	<ul style="list-style-type: none"> • An assessment unit with no data is between two assessment units that are Category 5 for the same pollutant
Other lines of evidence	<ul style="list-style-type: none"> • Documented fish kill • Studies or other data/info that demonstrate impairment

DEQ would use the following guidelines to assign assessment units to Category 3B:

Insufficient data	<ul style="list-style-type: none"> • At least 1 sample exceeds the magnitude of the criteria • AND dataset does not meet minimum size requirement for Category 5 • BUT no overwhelming evidence of impairment exists.
Conflicting indicators of attainment	<ul style="list-style-type: none"> • When samples measured as total recoverable exceed a dissolved criterion.
Data not quantifiable	<ul style="list-style-type: none"> • Exceeding samples below the method minimum reporting (MRL);
When assessing water-quality based criteria with defaults	<ul style="list-style-type: none"> • BOTH measured and default input criteria are used • AND measured input criteria sample data do not meet minimum sample size • AND some samples exceed criteria generated from default data

Alternative formats

Documents can be provided upon request in an alternate format for individuals with disabilities or in a language other than English for people with limited English skills. To request a document in another format or language, call DEQ in Portland at 503-229-5696, or toll-free in Oregon at 1-800-452-4011, ext. 5696; or email deqinfo@deq.state.or.us.