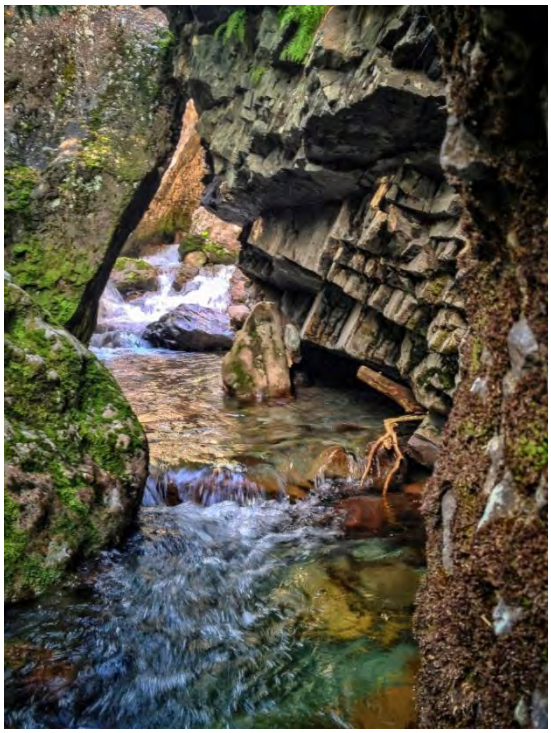
COLORADO
Department of Public Health & Environment

Section 303(d) Listing Methodology 2016 Listing Cycle



Colorado Department of Public Health and Environment
Water Quality Control Division

March 2015

SECTION 303(d) LISTING METHODOLOGY

2016 Listing Cycle

I. INTRODUCTION

Section 303(d) of the federal Clean Water Act requires states to identify waters where effluent limitations mandated by Section 301(b)(1)(A) and Section 301(b)(1)(B) are not stringent enough to attain water quality standards. These waters are compiled into the Section 303(d) list of impaired waters. The Colorado Section 303(d) List identifies those waterbodies where there are exceedances of water quality standards or non-attainment of uses. This includes waters impaired as a result of non-point source, point source discharges or combined point source and non-point source contributions including natural sources. Total Maximum Daily Loads (TMDLs) are required for each listed waterbody. The 2016 Section 303(d) List is equivalent to Category 5 waters in the Environmental Protection Agency's (EPA's) *July 29, 2005 Guidance for 2006 Assessment, Listing and Reporting Requirements Pursuant to Sections 303(d), 305(b) and 314 of the Clean Water Act* (2006 Integrated Reporting Guidance), and the *September 3, 2013 Information Concerning 2014 Clean Water Act Sections 303(d), 305(b) and 314 Integrated Reporting and Listing Decisions*.

The 2016 Monitoring and Evaluation List (M&E List) identifies waterbodies where there is reason to suspect water quality problems, but there is also uncertainty regarding one or more factors, such as the representative nature of the data. The M&E List contains waterbodies that would be reflected in Category 3 of EPA's Integrated Reporting Guidance.

Waters that are on neither the Section 303(d) List nor the M&E List are:

- Attaining their uses and standards (EPA's Category 1);
- Attaining some uses (EPA's Category 2);
- Have not been fully assessed (EPA's Category 2 or 3); or
- Impaired but do not require a TMDL for the following reasons (EPA's Category 4):
    - 4a - TMDLs have been completed but uses are not yet attained;
    - 4b - other required control mechanisms are expected to address all waterbody-pollutant combinations and will attain water quality standards in a reasonable period of time;
    - 4c - the impairment is not caused by a pollutant.

Section II of this document identifies the process that the Water Quality Control Division (division) and Water Quality Control Commission (commission) intend to follow in establishing the Section 303(d) and M&E Lists. Section III contains the listing criteria and Section IV contains the prioritization criteria.

This document provides a framework for the determination of attainment or non-attainment of assigned water quality standards and designated uses. However, there may be site specific considerations not identified in the listing methodology that are appropriately factored into the final listing decision. Generally, the division's recommendation to list or not list a waterbody will be based upon stringent application of the listing methodology criteria, but best professional judgment (BPJ) may be applied when necessary. Parties will have the opportunity to present mitigating evidence for the commission's consideration as part of the rulemaking hearing process.

II. LISTING PROCESS

A. Development of the Methodology

The listing methodology is reviewed and updated on a biennial basis in anticipation of 303(d) List and M&E List development. The listing methodology is revisited and revised with the intent of clarifying the division's procedures for assessing attainment of those uses and standards assigned by the commission to Colorado waters. Most often revisions or additions to the listing methodology derive from issues raised during the previous listing process.

The division has solicited public participation to develop the 2016 Section 303(d) Listing Methodology through several means. The methodology for development of the 2012 list was used as a starting point (however, the listing methodology began as the introduction to the 1998 lists, later becoming a standalone document in 2002). The listing methodology is developed in a public process. Work group meetings to develop the 2016 303(d) Listing Methodology were held on March 5, 2014; May 7, 2014; August 6, 2014; September 3, 2014; October 1, 2014; November 5, 2014 and December 2, 2014.

B. Process for Adopting the Methodology

The process for formal consideration and acceptance of the listing methodology was discussed at an April 2003 commission meeting. At that time, the commission decided to convene an Administrative Action Hearing (AAH) process for adoption of the listing methodology. Since the 2004 cycle, the listing methodology has been approved in an AAH process. The following schedule is anticipated for development and finalization of the 2016 Section 303(d) listing methodology:

- The division proposal will be available for public review by January 8, 2015 as an attachment to the notice for the March 10, 2015 public hearing on the listing methodology. The proposal will also be available on the commission's website, emailed to participants in the 303(d) Listing Methodology work group and the notice will be published in the monthly Water Quality Information Bulletin.
- The notice will establish a deadline of January 29, 2015 for written comments on the proposed listing methodology, including any recommendations for alternative language in the document. Comments received will be posted on the commission's web site and copies will be available in the commission office.
- The notice will also establish a deadline of February 19, 2015 for any written rebuttal comments in response to the January 29, 2015 comments. These rebuttal comments will be posted on the commission's web site and copies will be available in the commission office.
- If the initial written comments and/or the rebuttal comments warrant revisions to the proposed listing methodology, the division will submit a revised proposal by February 25th. This revised proposal will be posted on the commission's website and copies will be available in the commission office.
- No other written materials will be accepted for this hearing except by specific permission from the commission, with written explanation as to why such materials could not have been submitted in accordance with the above deadlines.

- There will be opportunity at the March 10, 2015 hearing for any interested persons to provide oral comments regarding the proposed listing methodology.
- At the conclusion of the March 10, 2015 Administrative Action Hearing, the commission will modify, as necessary, and approve the final 2016 Section 303(d) Listing Methodology.

C. Process for Adopting the Section 303(d) and Monitoring & Evaluation Lists

The process for formal consideration and adoption of the Section 303(d) and M&E Lists was also discussed at the April 2003 commission meeting. The commission decided that the 2004 lists, and subsequent lists, would be adopted through a public rulemaking process. The following steps will be used for the adoption of the 2016 303(d) and M&E Lists:

- Any person that has a Category 4b demonstration plan that wishes the division to consider and submit to EPA must provide that information to the division by the last week in August. The division will formally submit the plan to EPA by the first week in September.
- Any person that has data or other information that it wishes the division to consider in determining which water segments and parameters to propose for listing or delisting (for either the Section 303(d) List or the M&E List) must provide that information to the division by April 15, 2015. The division will formally notice its solicitation of data, with instructions for its format, for consideration in development of the 2016 Section 303(d) List in January 2015.
- By the third week in June 2015, external parties contact the division with suggestions for the 303(d) List and/or the Monitoring and Evaluation List.
- By the fourth week in June 2015, the division responds to the external parties regarding whether the segments in question will be in their proposal or not (giving external parties three weeks to develop their own proposal).
- Any person who wishes to propose the listing of water segments/parameters that may not be proposed by the division must submit any such proposal, with accompanying proposed statement of basis and purpose language, to the commission office by July 15, 2015. Any such proposal must also include adequate information for the commission to determine that listing of the segment/parameter should be considered in the rulemaking.
- The division will also submit its proposal to the commission by July 15, 2015.
- A draft rulemaking hearing notice, with the division's and any external proposals attached, will be prepared by the commission office, for inclusion in the commission's August meeting packets. The draft notice and proposals will also be posted on the commission's web site by the first week in August 2015, and will be emailed to the work group.
- The commission will review the draft notice and proposals at its August 10, 2015 regular meeting and approve them for filing.
- The rulemaking hearing notice and proposals will be filed with the Secretary of State by August 31, 2015. The final notice and proposals will also be posted on the commission's web site by about this date, and will be emailed to the 303(d) Listing Methodology work group.
- The rulemaking notice will include contact information for persons wishing to get more detailed information regarding data or other information supporting the listing proposals advanced by the division or other persons.

- The rulemaking hearing notice and proposal will be published in the September 2015 Colorado Register.
- The notice will establish a party status deadline of the last week of September, 2015.
- Written proponent's prehearing statements will be due by the first week of October, 2015.
- Responsive prehearing statements and any evidence (data and any other relevant information) regarding potential listing will be due by the last week in October.
- This October deadline for the submission of evidence (data and any other relevant information) will apply to any information from any interested persons, not just those with party status.
- A prehearing conference will be held during the second week of November, 2015.
- The notice will provide an opportunity for the submission of written rebuttal statements, in response to the October submissions, by the last week in November, 2015. No new data or other new factual information will be accepted after the last week of October but the rebuttal statements may contain different analyses and perspectives regarding what the submitted information shows regarding attainment and the appropriateness of listing and may include additional information solely to rebut or respond to information filed with another party's responsive prehearing statement.
- Any data or other information that is not submitted in accordance with the above deadlines will be considered in the next listing cycle.
- The commission's rulemaking hearing will be held on December 14, 2015. At the conclusion of the hearing, the commission will approve the 2016 Section 303(d) List and the Monitoring and Evaluation List as Regulation #93.

Table 1. Summary of dates for adoption of the 303(d) Listing Methodology and Regulation #93

TOPIC	IMPORTANT MILESTONE	APPROXIMATE DATE
303(d) LISTING METHODOLOGY	Draft proposal deadline	1/8/2015
	Written comments due	1/29/2015
	Rebuttal comments due	2/19/2015
	Revised proposal due	2/25/2015
	Administrative Action Hearing (AAH)	3/10/2015
CATEGORY 4b	Category 4b Plan due to division	Last week in Aug. 2015
	4b Plan due to EPA	1st week in Sept. 2015
DATA CALL	Data call	1st week in Jan. 2015
	Data submittal due	4/15/2015
REGULATION #93 303(d) and M&E LIST	External suggestions for list	3rd week in June 2015
	Division response to suggestions	4th week in June 2015
	Division and third party proposals due	July 15, 2015
	Draft rulemaking hearing notice	1st week in Aug 2015
	Proposal review at commission meeting	8/10/2015
	Rulemaking hearing, notice and proposal filed with Secretary of State	8/31/2015
	Hearing notice and proposal published in register	9/10/2015

Table 1. Summary of dates for adoption of the 303(d) Listing Methodology and Regulation #93		
TOPIC	**IMPORTANT MILESTONE**	**APPROXIMATE DATE**
	Deadline to establish party status	4th week in Sept. 2015
	Written proponents prehearing statements due	1st week in Oct. 2015
	Responsive prehearing statements and evidence due	Last week in Oct. 2015
	Prehearing conference	2nd week in Nov. 2015
	Rebuttal comments due	Last week in Nov. 2015
	Rulemaking hearing for 303(d) List and M&E List	12/14/2015

D. Process for Removing Waterbodies from the Section 303(d) and Monitoring and Evaluation Lists

This document addresses the procedures and protocols utilized by the division in assessing information for the purpose of identifying instances of non-attainment of water quality standards and subsequently, inclusion of affected waterbodies on either the 303(d) or M&E List. In general, removal of waterbodies/pollutants from either list is subject to the same requirements as those utilized for addition to the lists. Removal from the lists is considered appropriate where new information is developed which indicates that water quality standards are being met and/or designated uses are being attained. Considerations include more recent or more accurate data (for instance, chemical data generated using clean sampling/analytical methods), more sophisticated analysis using a calibrated model, identification of deficiencies in the original assessment or changes in standards, guidance or policy.

Sampling frequency and number of sampling events needed to delist (or remove) a waterbody should be similar to, or greater than, that which was used as a basis to list the segment (an exception would be in the instance where data collected utilizing conventional methods is supplanted by clean methods data or where the listing decision was based upon special study results for which it is impractical to reproduce). In any case, data must be adequate to characterize current water quality conditions. Assessments demonstrating attainment of designated uses should provide documentation of a nature similar to that used to support the listing decision. Attainment of water quality standards and uses will result in removal of the waterbody, or one or more listed parameters, from the list.

The commission will also consider removal when *good cause* is shown. As described in EPA's **2006 *Integrated Reporting Guidance*, *good cause*** for removing a waterbody (or water body pollutant combination) from the lists includes:

- The assessment and interpretation of more recent or more accurate data in the record demonstrate that the applicable classified uses and numeric and narrative standards are being met.
- The results of more sophisticated water quality modeling demonstrate that the applicable classified uses and numeric and narrative standards are being met.

- Flaws in the original analysis of data and information led to the waterbody pollutant combination being incorrectly listed.
- Demonstration pursuant to 40 CFR 130.7(b)(1)(ii) that there are effluent limitations required by state or local authorities that are more stringent than technology based effluent limitations, required by the Clean Water Act, and that these more stringent effluent limitations will result in attainment of classified uses and numeric and narrative standards for the pollutant causing the impairment.
- Demonstration that there are other pollution control requirements required by state, local or federal authorities that will result in attainment of classified uses and numeric and narrative standards within a reasonable time. (This element is EPA's Integrated Reporting Category 4b.)
- Documentation that the state included on a previous Section 303(d) List an impaired water that was not required to be listed by EPA regulation, e.g. waters where there is no pollutant associated with the impairment (This element is EPA's Integrated Reporting Category 4c).
- Approval or establishment by EPA of a TMDL since the last Section 303(d) List. (This element is EPA's Integrated Reporting Category 4a.)
- Inappropriate listing of a water that is located within Indian lands as defined in U.S.C. § 1151: *Indian Country Defined*.
- Other relevant information that supports the decision not to include the segments on the Section 303(d) List:
    - Adoption of revised water quality standards and/or uses such that the water is now in attainment of the revised standards and/or uses;
    - Development of a new listing methodology consistent with the state water quality standards and classifications and federal listing requirements;
    - A reassessment of the data that led to the prior listing, concluding that the waterbody is no longer impaired.

Barring unforeseen circumstances, the division will only propose to revise the lists during the regularly scheduled reviews (currently biennially). Other interested persons may petition the commission at any time to request a rulemaking hearing to revise the lists (either additions or deletions). However, such a hearing will be held only upon showing that failing to either add a segment to the list or delete a segment from the list prior to the next scheduled review will result in a substantial hardship to the party or parties requesting the hearing.

E. Process for Determining Category 4b Classification

An alternative to listing an impaired segment on the state's 303(d) List is an approved Category 4b demonstration plan. A Category 4b demonstration plan, when implemented, must ensure attainment with all applicable water quality standards through agreed upon pollution control mechanisms within a reasonable time period. These pollution control mechanisms can include approved compliance schedules for capital improvements or plans enforceable under other environmental statutes (such as the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA)) and associated regulations. A Category 4b demonstration can be used for segments impaired by point sources and/or non-point sources. Both the division and EPA must accept a Category 4b demonstration plan for the affected segment to be

placed in Category 4b. In the event that the Category 4b demonstration plan is not accepted, the segment at issue will be included on the 303(d) List, Category 5.

Generally speaking, the following factors will be considered necessary for Category 4b demonstration plan acceptance: (1) appropriate voluntary, regulatory or legal authority to implement the proposed control mechanisms (through permits, grants, compliance orders for Colorado Discharge Permit System permits, etc.); (2) existing commitments by the proponent(s) to implement controls; (3) adequate funding; and (4) other relevant factors appropriate to the segment.

The following evidence must be provided as a rationale for a Category 4b demonstration plan:

1) A statement of the problem causing the impairment;
2) A description of:
    a. pollution controls to be used;
    b. how these pollution controls will achieve attainment with all applicable water quality standards;
    c. requirements under which those pollution controls will be implemented;
3) An estimate of the time needed to meet all applicable water quality standards;
4) A schedule for implementation of the necessary pollution controls;
5) A schedule for tracking progress including a description of milestones; and
6) A commitment from the demonstration plan proponent to revise the implementation strategy and pollution controls if progress towards meeting all applicable water quality standards is not shown.

Timing for proposal submittal and acceptance by the division and EPA:

- Category 4b demonstration plans should be submitted to the division by the last week in August, 2015 in order for the division to submit the plan to EPA by the first week in September, 2015. Parties are encouraged to work with the division well in advance of this date as states are the entity required to submit these plans to EPA.

- Acceptance from EPA must be obtained by the last week in October, 2015 otherwise the division will continue to propose that the segment in question is included on the 2016 303(d) List.

- If EPA and the division accept the Category 4b Plan, the division will notify the commission and public through proposed statement of basis and purpose language in its proposal that a Category 4b demonstration plan is accepted and is appropriate for this segment.

- Category 4b demonstration plans must be included in either the proponents prehearing statement or the party's responsive prehearing statement.

- Category 4b segments will be included in Regulation #93, in a table following the 303(d) List and will be reported to the EPA as a part of the Integrated Report.

EPA has several documents that contain additional information on Category 4b demonstration requirements, including the *2006 Integrated Report Guidance*, available at http://www.epa.gov/OWOW/tmdl/2006IRG/#documents; and *Information Concerning 2008 Clean Water Act Sections 303(d), 305(b), and 314 Integrated Reporting and Listing Decisions*, available at: http://water.epa.gov/lawsregs/lawsguidance/cwa/tmdl/2008_ir_memorandum.cfm

F. Process for Determining Category 4c Classification

In cases where the impairment is determined to be caused exclusively by pollution, that does not result in pollutant(s) levels in excess of state water quality standards, the impaired waterbody may be placed into Category 4c. As defined by the Clean Water Act, pollution is "the man-made or man-induced alteration of the chemical, physical, biological and radiological integrity of water" whereas pollutants are "dredged spoil, solid waste, incinerator residue, sewage, garbage, sewage sludge, munitions, chemical wastes, biological materials, radioactive materials, heat, wrecked or discarded equipment, rock, sand, cellar dirt and industrial, municipal, and agricultural waste discharged into water." (section 502(19) and (6)). Segments classified as Category 4c are impaired however, a TMDL will generally not be required. Examples of circumstances where an impaired waterbody segment may be placed into Category 4c include segments impaired solely due to lack of adequate flow or to stream channelization. While low flows may be a human-induced condition (i.e., a reduced volume of water) fitting the definition of pollution, lack of flow sometimes leads to the increase of the concentration of a pollutant (e.g., sediment) in a water, such that a TMDL, which may consider variations in flow, is required. Segments located below dams or stream diversions with impaired biological communities (indicated by a failing multimetric index (MMI score) not caused by the presence of a pollutant(s), may be candidates for Category 4c.

Before placing impaired waterbody segments into Category 4c, thorough monitoring and assessment needs to be performed on the segment to confirm that no pollutants are contributing to the waterbody's failure to meet water quality standards. If adequate monitoring and assessment is not performed to rule out pollutant(s) as a cause, then the impaired waterbody should be placed on the 303(d) List (Category 5).

Proposals for Category 4c should be submitted as a part of the 303(d) rulemaking process in the division's proposal or a proponent's proposal, which is attached to the notice of rulemaking. Documentation of pollutant investigations such as chemistry data, proof of impairment and support of the identified pollution source must be submitted as part of the proponent's prehearing statement. Category 4c segments will be included in Regulation #93, in a table following the 303(d) List and will be reported to the EPA as a part of the Integrated Report.

III. LISTING CRITERIA

This listing methodology sets forth criteria that generally will be used to make decisions regarding which waters to include on the 2016 Section 303(d) List and the 2016 M&E List. However, this methodology is not adopted by the commission as a rule. The commission is not bound by the criteria set forth in the listing methodology in making individual listing decisions if they determine on a site specific basis that an alternative approach provides a more

appropriate method for assessing attainment of water quality classifications and standards in a particular circumstance.

A. Existing and Readily Available Data

In determining whether data and information is existing and readily available, the division will take into account such data and information as it has utilized in the preparation of those identification processes, calculations, and models referenced in 40 CFR §130.7(b)(5)(i), (ii) and (iv) and that credible data and information presented in a readily usable format and submitted in reports provided to the division as referenced in 40 CFR §130.7(a)(5)(iii). In addition, the division will accept and take into consideration credible data and information that is submitted to the division as part of the listing process, within specified data submittal deadlines. The division will also continue to independently collect and analyze new data on a rotating basin basis as part of its triennial review efforts and will utilize such data and information in making listing determinations. Existing data, which are not brought forward through one of the above mechanisms or otherwise presented to the division in accordance with the schedule, set out in Section II.C, above, will not be treated as readily available for purposes of making listing decisions. Such information will be considered in the next listing cycle.

It is important that data submitted for consideration in the 303(d) List development process is in a form that is amenable to existing division data management capabilities. Chemical data submitted for consideration in the list development process should be submitted in an electronic, WQX-compatible format. Physical and biological data should be submitted in a common electronic format that can be analyzed statistically. Recommended data reporting templates will be available at the time of the data request in January 2015. The division must be consulted regarding alternate formats. Data that are submitted in hard copy or alternate electronic format will be considered subject to division resource limitations, and may not be included in the division's assessment or proposal.

The assessment process is intended to provide continuity with similar assessments done to support the standards review process as well as to efficiently utilize division resources. The division uses a rotating basin approach, approved by EPA, for periodic standards review and coordinates water quality monitoring and assessment to support the review. The following schedule sets out the relationship between basin reviews and when assessments generated by those reviews will be incorporated in the 303(d) listing process for the first time.

Table 2. Coordination between the Standards Review Schedule and Section 303(d) List Cycle				
River Basins (Regulation Number)	**Data Collection Effort**	**Data Call/ Assessment Season**	**Standards Hearing**	**Assessments Incorporated into 303(d) List**
Arkansas & Rio Grande (#32 & #36)	July 2010 – June 2011	Spring 2012	June 2013	(2014 cancelled) Dec. 2015
Colorado Basin (#33 & #37)	July 2011 – June 2012	Spring 2013	June 2014	(2014 cancelled) Dec. 2015
South Platte (#38)	July 2012 - June 2013	Spring 2014	June 2015	Dec. 2015
Basic Standards (#31)	July 3013 – June 2014	Spring 2015*	June 2016	Dec. 2015
San Juan, Dolores & Gunnison (#34 & #35)	July 2014 - June 2015	Spring 2016	June 2017	Dec. 2017
Arkansas & Rio Grande (#32 & #36)	July 2015 - June 2016	Spring 2017	June 2018	Dec. 2017
Colorado Basin (#33 & #37)	July 2016 - June 2017	Spring 2018	June 2019	Dec. 2019
South Platte (#38)	July 2017 - June 2018	Spring 2019	June 2020	Dec. 2019

* Statewide data call for additional data for Regulation #93 and the 2016 303(d) and M&E Lists.

B. Credible Evidence

The water quality assessment process depends on analysis of sufficient reliable data. Listing decisions not supported by adequate data are potentially flawed. The listing criteria are intended to assure that only those waterbodies for which adequate documentation of non-attainment is available are included on the Section 303(d) List. Waterbodies for which there is evidence to suggest impairment, but for which such documentation does not meet the standards for credible evidence, will be placed on the M&E List unless good cause is shown that it should be included on the 303(d) List.

Waterbodies may be included on the Section 303(d) List based on an evaluation of biological, chemical or physical data. The division will consider proposing to list a waterbody based upon consideration of all chemical, physical, and biological information that meets established sampling, analytical and interpretive protocols. Considerations include a review of the sampling and analytical methods employed. Factors to be considered include analytical detection limits, sample size (see section III.D.5.e), spatial and temporal distribution (see section III.C.1.f), variability within the data set, and the use of clean methods. Listing is often based upon chemical data alone, subject to the data interpretation criteria identified within this document. Listing based upon biological or physical data in the absence of accompanying chemical data requires that such information clearly demonstrate use impairment. Only representative data will utilized as the basis for the listing decisions.

The following guidelines are used to evaluate the adequacy of water quality information as a basis to support listing a waterbody.

1. Data Requirements – General

Information must be available to describe the methods used for sample collection, field and laboratory analysis. Persons submitting data for consideration during the list development process must either provide the relevant quality assurance documentation with the submittal or assure that the documentation is available for the division to review.

The party submitting the data for consideration should provide the following information accompanying their data submission:

- Written assurance that the methods and procedures specified in the Quality Assurance (QA) plan were followed.

- Any field notes, laboratory comments, or laboratory notations concerning a deviation from standard procedures, quality control, or quality assurance that affects data reliability, data interpretation, or data validity may be requested by the division.

- Statement of the analytical methods used by the laboratory, method number, detection limits, quantitation or minimum levels, if available and any quality control samples and standards necessary to properly interpret data different from that stated in the QA plan.

- If requested by the division for interpreting or validating data, any other information, such as complete field notes, photographs, climate, or other information related to flow, field conditions, etc. This information should be retained by the submitter for a period of at least five years.

- Field instruments, such as multi-parameter devices, must be operated and calibrated according to manufacturer's recommendations or other acceptable demonstrated method. Calibration information and any other documentation of accuracy may be requested by the division.

Minimum information required for each data submittal must include the following:

- Location of each sample station in latitude and longitude with the associated reference datum, e.g., North American Datum 1983, etc.
- Waterbody name and sampling location description.
- Date the sample was taken.
- Parameter or condition measured.
- Measured value.
- Unit of measurement.
- For non-detect or non-quantifiable data, the less than value associated with the method detection limits (MDL) or reporting limits (RL) (ie., LQL, LRL, PQL, etc.).
- Method used to measure the pollutant.
- Name and contact information of the party submitting the data.

Data submittals must include precise, sufficient information on the name of the waterbody and location of the sample station to allow for accurate mapping.

2. Sampling and Analysis Plans

Chemical data should be supported by a Sampling and Analysis Plan (SAP), which identifies sampling locations, contains analytical method references, and incorporates Quality Assurance/Quality Control (QA/QC) provisions. QA/QC documentation may include references to a standard QA/QC protocol. During review of chemical data submitted for evaluation, the division may require submittal of the SAP, QA/QC protocols and the results of QA/QC efforts. The division will provide any such information to other parties upon request.

3. Toxicity Tests

In-situ bioassay test results, or other ambient toxicity test results, must demonstrate adverse effects as measured by a statistically significant response relative to a representative reference or control. Inherent variability in toxicity testing results must be adequately taken into account. Listing decisions based upon toxicity test results require that any such results be corroborated by biological information clearly demonstrating impacts to aquatic community health, composition, or productivity. Data received utilizing whole effluent toxicity (WET) methods will be considered on a case by case basis.

4. Physical and Biological Assessment

a. Physical and biological assessments must be performed in accordance with scientifically sound methodologies. All such assessments should be performed by an observer who has training and experience in performing such evaluations. Assessment reports should include a statement of the observer's qualifications and should reference the protocols utilized. Any departures from referenced protocols and methodologies should be documented and the basis for any such departure addressed. The division's recommended collection and assessment methodologies for physical and biological data are described in commission Policy 98-1, *Guidance for Implementation of Colorado's Narrative Sediment Standard in Regulation #31, Section 31.11(1)(a)(i)* and Policy 10-1, *Aquatic Life Use Attainment, Methodology to Determine Use Attainment for Rivers and Streams.* A description of the division's assessment procedures for both sediment and aquatic life are ~~is~~ included in section III.D.7.a & d.

b. The division will generally accept methodologies and protocols in use by the U.S. Geological Survey, U.S. Forest Service, U.S. Bureau of Land Management, EPA, Colorado Parks and Wildlife, or others, when well documented, widely available and suitable for their intended purpose. The division's determination of the acceptability or unacceptability of any such protocol will be included in the division's discussion of data sources included in the pre-hearing statement of the Section 303(d) List.

5. Period of Record

The division will announce a request for data (data call) in January of each year. Data will be received by the division for approximately three months. Data and other information received will be assessed for the next water quality assessment session. The assessment results will be available for public review in either the basin regulation or the Regulation #93 review processes.

Data collected within the dates specified in the data call may be submitted for consideration into the assessment. In general, information and data should be no older than five years. Data which are less than five years old and meet the other credible data requirements outlined in this listing methodology will be consolidated and assessed with other data.

Data older than five years must meet all current data requirements and will only be considered on a case by case basis for the following reasons:

- No newer data exists for the waterbody segment/parameter or the existing data does not meet the requirements of this listing methodology;
- The data are part of a larger dataset or long term monitoring which includes data younger than five years old for the same waterbody/parameter;
- Information or rationale is provided with the data to show that the data reflects current conditions and adheres to acceptable protocols.

Data older than five years may be used when necessary to determine historical natural conditions if the data meets the QA requirements in place at the time of its collection.

Data submitted after the three month deadline stated in the data call will not be considered for the current assessment/listing cycle unless it is submitted in written testimony during the hearing process. Any other late data, accepted after the data request deadline, will be put into consideration for the next assessment/listing cycle.

6. Anecdotal Information

Anecdotal information, in the absence of chemical, physical, or biological data, will not in and of itself be adequate to support a listing decision unless such information provides clear and convincing evidence demonstrating non-attainment. Anecdotal information includes, but is not limited to fishing logs, field logs, and historical or archival documents.

7. Representative data

Representative data is defined on page 33, Section III.D.5.a. Non-representative data includes data collected within the mixing zone of a discharge. Data collected during or immediately after temporary events influencing the waterbody that are not representative of normal conditions shall typically be discounted in making the listing decision. For example, scouring storm flows which lead to diminished aquatic life use or accidental

spills of toxic chemicals would not be a basis upon which to list the affected segment. However, such events may be considered as a basis for listing in instances where non-attainment of standards arises from a reversible source of pollutants.

C. Data Interpretation

The water quality assessment process considers the numeric and narrative standards assigned to a segment, as well as the assigned use classifications. Numeric standards are identified for a given pollutant and are expressed as a threshold value or as an acceptable range of values. Determination of attainment/non-attainment of pollutant specific numeric standards is a relatively straightforward statistical process.

Narrative standards describe threshold conditions that, if exceeded, result in unacceptable water quality conditions. Narrative standards that are applied to all surface waters in Colorado address sediment, floatables, film, odor, taste, color, toxins and excessive nutrients. Narrative standards may also include site specific temperature standards as provided at section 31.7(1)(c) of the *Basic Standards,* Regulation #31. Exceedance of narrative standards is more difficult to ascertain, as there are typically no quantifiable expressions of parameter concentration or loading that result in non-attainment. It is often the impact of pollution or of a pollutant, and not the pollutant itself, which is observed.

Use classifications identify existing or potential uses of the surface water segment. These include aquatic life, water supply, recreation and agricultural uses. Specific numeric standards are attached to a given use classification. Assignment of an aquatic life use classification to a segment typically results in assignment of a related suite of numeric standards. Attainment of numeric standards serves as a surrogate measure indicating attainment of the assigned use classification. However non-attainment of an assigned use classification, as with narrative standards, may result from causes or parameters other than those assigned numeric standards.

1. Chemical Data - General

The division determines attainment of numeric standards by assessing data against applicable standards adopted for each waterbody segment. Additional methods used for the assessment of lakes and reservoirs are also noted.

a. <u>Attainment of Chronic Standards</u>

Attainment of chronic chemical standards, in both streams and rivers, and lakes and reservoir systems, is based upon the 85th percentile of the ranked data, except as otherwise noted below. Percentile values are calculated by ranking individual data points in order of magnitude. Hardness based metal standards are evaluated by comparing the 85th percentile against the assigned hardness based equation using mean hardness. Total recoverable metals are evaluated against the median value or the 50th percentile. Dissolved metals are evaluated against the 85th percentile. Dissolved oxygen (DO) is evaluated at the 15th percentile for streams. Minima pH is evaluated against the 15th percentile, maxima at the 85th percentile.

Hardness based metals standards may also be evaluated by a detailed assessment where the chronic table value standard is calculated for each paired hardness/concentration data. A waterbody is considered impaired if the standard is exceeded more than 15% or 50% of the time for dissolved and total recoverable metals, respectively. In the case where both the average hardness and paired hardness assessment approaches are conducted and the listing decision differs between the two approaches, the paired hardness/concentration assessment decision is considered more representative.

b. Attainment of Acute Standards

Acute standards are evaluated by comparison of single sample values to the assigned standard. The acute table value standard is calculated for each paired hardness/concentration and attainment is determined for each data pair. If more than one data pair exceeds the standard within a three year period, the waterbody is considered impaired. Where paired hardness and concentration data is not available, an assessment of the acute standards cannot be completed.

c. Spawning Season DO Criteria

Spawning season DO criteria, when assigned by the commission, are generally applied for the period between mid-October through July (dependent upon species present and basin). Attainment of the spawning season DO standard is evaluated through a two step process. An initial screening is performed by comparison of the 15th percentile DO value to the 7.0 mg/L spawning season based standard. In instances where the 15th percentile value for the entire dataset is less than the 7.0 mg/L seasonal standard, the dataset is subdivided into spawning/non-spawning values and the 15th percentile value for the spawning season data is compared to the spawning season criterion. The division will generally utilize spawning season information as provided by Colorado Parks and Wildlife on a basin specific basis (see Appendix A). Where more detailed fishery community information is available, the division will consider alternate spawning seasons as supported by such data.

d. Detection Limits

Sample data that are below detection limits will, in general (except *E. coli* data), be treated as zeros for assessment of attainment.

e. *E.coli* Standards

Attainment of the *E.coli* standard is assessed using the geometric mean of representative stream samples. Notwithstanding the criterion at item d above, *E.coli* data that are reported as less than detect will be treated as a value of one to allow calculation of a geometric mean.

f. Sample Bias

Assessment techniques will be used that seek to reduce the effects of biased sampling. For example, the median of multiple samples taken within a seven day period will be used to represent that time period, and information gathered during synoptic (sampling at many locations at the same time) sampling events may be considered in a separate assessment so as not to bias the conclusions. Water quality data may be evaluated differently on a case by case basis if it is determined that data within a seven day period may not be representative of the given seven day sample period.

2. Sampling Data – Lakes and Reservoirs

The sampling strategy for lakes differs from streams in important ways that affect the assessment of water quality. Typically, lakes are not sampled as often as streams because the large volume of water buffers against the short term changes in quality. In contrast, the spatial coverage within the vertical profile should generally be more comprehensive for lakes, especially where lakes are stratified in the summer.

Typically, two strategies are applied simultaneously when sampling lakes – vertical profiles and discrete samples. It is common to measure some parameters, usually temperature, DO, pH and conductance, in vertical profiles that yield measurements at closely spaced intervals from top to bottom in a lake or reservoir. Profile data are essential for defining the internal boundaries of layers that form when thermal stratification develops in the summer months, or for demonstrating that no stratification exists. The preferred sampling location for lake and reservoir profiles is in the deeper part of the lake or reservoir, most commonly in front of the dam for a reservoir.

Profile Data

The interpretation of profile data begins by examining the mixed layer, which is that part of a lake that is well mixed by wind action and can be expected to have relatively homogenous physical and chemical conditions. For assessment purposes, the mixed layer is evaluated by examining conditions in the upper portion of a lake. The upper portion is generally characterized within a single profile as follows:

1. Where a lake or reservoir is equal to or greater than five meters deep, measurements within a single profile are generally assessed as the average of all measurements from 0.5 meter to 2.0 meters.
2. Where a lake or reservoir is less than five meters deep, but more than 1.25 meters deep, measurements within a single profile are generally assessed as the average of all measurements from 0.5 meter to a depth equal to 40 percent of total depth.
3. Where a lake or reservoir is 1.25 meters deep or less, measurements within a single profile are generally assessed as the median of all measurements.

In a stratified lake, the upper portion is separated from a deeper, cooler layer (referred to as the lower portion) by a transition zone of rapid temperature change (thermocline).

The lower portion of a lake is assessed by averaging the measurements from one to three meters above the bottom of the lake. For example, to assess the lower portion of a lake with a maximum depth of six meters, profile measurements would be averaged from three to five meters. For lakes less than five meters deep, the lower portion is not assessed. This definition for the lower portion of a lake is only used for the purpose of pH assessment.

In cases where multiple data points along a profile are not available or feasible and only single data points are collected, a single data point from each of the

upper and lower portions may be assessed against the standard, if the single data points are determined to be representative.

Discrete Samples

Discrete samples are used to characterize conditions at specific depths, often intended to represent a single layer. Discrete samples from lakes are analogous to grab samples taken for stream assessments. It is common to take samples from the top and bottom of each lake (which would correspond to upper and lower portions in a stratified lake) because the water quality characteristics of those two major habitat regions often diverge significantly during the growing season. It is much less common to take discrete samples from the thermocline for two reasons: it is a boundary zone with steep environmental gradients and water quality characteristics will be intermediate between those of the adjacent layers.

a. Temperature

Vertical profiles provide a record of temperature at closely spaced intervals from the top to the bottom of the water column. Unlike streams, daily fluctuation of temperature in lakes tends to be quite small. Thus, the temperature observed at each depth in the profile is assumed to be persistent on a scale of days, making it a surrogate for the weekly average temperature (WAT). The division examines attainment first in the upper portion, as defined above to represent the mixed layer referenced in Regulation #31 (footnote 5.c.iii). If the average temperature in this upper portion exceeds the chronic standard, temperatures below the upper portion are evaluated when there is need to consider adequate refuge. Adequate refuge depends on concurrent attainment within a given depth of the temperature standard and applicable dissolved oxygen standards (Regulation #31, Table 1, (footnote 5.c.iii)).

b. Dissolved Oxygen

Assessment of dissolved oxygen within a profile of a lake or reservoir is accomplished by comparing the average of the measurements from the upper portion of the lake, as defined above, to the applicable standard. Dissolved oxygen is not assessed in the lower portion (bottom layer) of the lake, except where adequate refuge is necessary for assessment of the temperature standard.

Fall turnover exclusion: DO may drop 1 mg/l below the criteria in the upper portion of a lake or reservoir for up to seven consecutive days during fall turnover provided that profile measurements are taken at a consistent location within the lake seven days before, and seven days after the profile with low DO. The profile measurements taken before and after the profile with low DO must attain the criteria in Table 1 (Regulation #31) in the upper portion of the lake or reservoir. The fall turnover exclusion does not apply to lakes or reservoirs with fish species that spawn in the fall unless there are data to show that adequate DO is maintained in all spawning areas for the entire duration of fall turnover (Regulation #31, Table I, Footnote 9(e)(i)).

c. pH

Data for pH often are available from vertical profiles, but the data are generally evaluated in the context of discrete samples. There are two reasons for this

approach – not all sampling programs obtain pH in profiles, and pH must be determined for any discrete sample wherever ammonia is of interest. Discrete samples from the upper and lower portions are evaluated separately because they represent different habitat regions in a stratified lake. When variations in pH are driven largely by biological processes within a lake, the risks of exceedance are generally associated with high pH in the upper portion (due to high rates of algal productivity) and low pH in the lower portion (due to high rates of decomposition).

Assessment of the pH standard for a lake is accomplished by calculating the average pH from the upper and lower portions of the lake for each profile as defined above in the profile data section (III.2). The 15th and 85th percentiles of the sample averages for each portion are then compared to the minima and maximum pH standard for the determination of attainment.

d. Metals and Inorganics

These constituents are generally assessed on the basis of discrete samples (grab samples), for which the methods for data interpretation have been outlined above in the discrete samples section (III.2). For the reasons explained under the subsection on pH, discrete samples from the upper and lower portion of a lake should be assessed separately.

3. Biological and Physical Data

Biological and/or physical assessment protocols may support a determination of nonattainment of numeric standards or, alternately, nonattainment of narrative standards and classified uses. The division, in interpreting physical and biological information, will give site specific consideration to the applicability of the protocols in use and available metadata gathered to validate the information generated, the extent and nature of expertise of the observer and the relative weight of the evidence presented.

In general, a determination that an assigned aquatic life use is not supported will be consistent with the protocols established in commission Policy 10-1, *Aquatic Life Use Attainment, Methodology to Determine Use Attainment for Streams and Rivers.*

Physical and biological assessments will consider measurable conditions or features within an affected segment in comparison to an expected condition. The expected condition generally will be based upon a selected reference condition. Identification of reference conditions requires consideration of the following:

- level of disturbance (minimal)
- location (upstream, downstream or within a separate drainage)
- historical condition
- expected condition based on modeling or general expectations for highly managed systems
- other fair and reasonable comparisons

Determination of reference conditions based upon sampling/assessment of multiple reference sites, when possible, is preferable but not required.

Impairment of aquatic life use classifications or narrative toxicity standards will be demonstrated, for the limited purpose of listing, when either the physical/habitat data or biological community metrics reflect a condition that is significantly less than reference condition, or as outlined in Policy 10-1.

The division's assessment process includes documentation of data sources, an evaluation of the validity of the data, the appropriateness of the methodologies utilized and whether the data are representative. This latter element considers spatial and temporal variability in the dataset, as well as the age of the data, any relevant changes within the watershed that might affect the interpretation of the dataset, and any bias which might be inherent in the sampling plan. If the dataset for the affected reach is found to be representative and valid, a comparison is made to an expected condition. The expected condition may be defined by actual conditions upstream, or downstream of the affected reach, or may be defined by a comparable reach located in a differing drainage or watershed. Alternately, the expected condition may be developed based upon biocriteria, modeling or professional judgment. Any assessment must describe the basis for defining the expected condition.

When an assigned aquatic life use is determined to be impaired, listing will be made consistent with the procedures outlined in Section D.1, Determination of Impairment.

In instances where aquatic life use impairment is demonstrated by biological assessment, the data will be interpreted, when appropriate, as outlined in the commission Policy 10-1, *Aquatic Life Use Attainment- Methodology to Determine Use Attainment for Rivers and Streams.*

Consistent with Policy 10-1, locations of macroinvertebrate sample collections will determine the biotype assignment for the MMI assessments. Where uncertainty exists regarding the transitional boundaries between biotypes, the MMI for the adjacent biotype may be used to help determine the status of the aquatic life use. This additional analysis may be conducted under two circumstances:

1. At sites in Level IV Ecoregion 21c where the biotype assignment along a waterbody varies between biotypes 1 and 2 because the stream slope fluctuates above and below 0.04. This situation typically occurs when stream slopes are slightly greater than or less than 0.04 along the gradient of a waterbody resulting in varying site classifications or biotypes.

2. At sites that encompass the physical border between two different Level IV Ecoregions or elevation zone boundaries used in the biotype classification. This results in a predicted site classification in one biotype, but is narrowly adjacent to another biotype. In such cases, sites may be represented by characteristics shared by more than one biotype.

For these circumstances, MMIs for each of the adjacent biotypes shall be investigated and used in the assessment.

In instances where aquatic life use impairment is the result of excessive sediment deposition, the interpretation of such data will be outlined in the commission Policy 98-1, *Guidance for Implementation of Colorado's Narrative Sediment Standard, Regulation #31, Section 31.11, (1)(a)(i)*. The division will review sediment data and MMI scores and make listing decisions for each independently based on each policy document.

4. Outstanding Waters

Attainment of water quality standards assigned for those segments designated as outstanding waters will be based upon the evaluation of ambient water quality characteristics and biological /physical data as described in the preceding paragraphs C.1, C.2 and C.3. Attainment or maintenance and protection at their existing levels is assessed by comparison of current ambient water quality against water quality conditions at the time of designation (See ***Basic Standards and Methodologies for Surface Water***, 5 CCR 1002-31, section 31.8(1)(a)). The time of designation can usually be found in the statement of basis and statutory basis in the basin regulation for the segment in question.

5. Assessment of Temperature Data

Numerical temperature standards are evaluated against representative instream data. Temperature varies within a reach both spatially and temporally, e.g. summer and winter. Data should be taken from a location in the stream that is representative of the reach at the time the data are collected. For example, data should not be relied upon that are taken only in locations that may be substantially warmer or cooler than the rest of the segment (e.g. backwater habitats, eddies, deep pools or refugia).

a. Seasonal Maxima

i. Chronic: Attainment of the chronic numeric temperature standard is based upon a maximum weekly average temperature (MWAT), unless otherwise specified in a site specific standard. The MWAT is defined as an implementation statistic that is calculated from field monitoring data. The MWAT is calculated as the maximum weekly average temperature (WAT). The WAT is a simple moving average (rolling average) that uses a minimum of three equally spaced measurements throughout a 24 hour period over a seven day consecutive period. MWAT are not to be overlapped (i.e. temperature data used in the calculation of one exceedance of an MWAT will not be used in any other exceedance calculation).

ii. Acute: Attainment of the acute numeric temperature standard is based upon a daily maximum (DM) water temperature, unless otherwise specified in a site specific standard. The DM is defined as the highest two hour average water temperature recorded during a given 24 hour period.

b. Lakes and Reservoirs:
For lakes and reservoirs the WAT is assumed to be equivalent to the average temperature of the upper portion. As an initial screen, the upper portion is assessed using the average temperature of the top 0.5 – 2 meters as defined above for shallow lakes. When, upon this initial screen, the average temperature in the upper portion exceeds the chronic temperature standard, the division will analyze the available data for adequate refugia.

6. Portioning of Segments

Initially, all data submitted to the division for 303(d) assessment is evaluated by sampling location. Subsequently, data from each station within a segment are combined for assessment of the segment as a whole. If data is only available for a limited area of a segment, then the conclusion reached from that area will be applied to the entire segment, if the sampling area is representative of conditions that exist within the entire segment. For segments that indicate non-attainment, the division will investigate further to determine whether the impairment is widespread or limited to individual portions of the segment such as specific tributaries or reaches. Supplemental information will be considered when determining the geographic extent of impairments. This information could include but is not limited to chemistry data, landscape analysis (i.e. hydrology, vegetation, soils, and elevation), underlying geology and an investigation of activities in the watershed (see Table 3).

Table 3. Listing Decisions based on Preliminary Source Information

Preliminary Source Information	**Listing Decision**
No discernible source information	List entire segment
Geogenic source (example: selenium from shale formations)	List portion that shares common geogenic source formations
Suspected point or area source (example: metals from a legacy mining feature)	List portion effected by the suspected point or area source
Sediment from anthropogenic sources or erosion (example: dirt road crossings of streams)	List portion effected by the source causing sediment
Temperature impacts from diversions and dams	List portion effected by the diversion or dam
E. coli or pathogens in urban areas (example: tribs to South Platte in Denver)	List entire segment
E. coli or pathogens in rural areas	If no upstream source is suspected (CAFO, septic systems), then list entire segment. If an upstream source is suspected, list only portion effected.

D. Determination of Impairment

Application of chemical, physical and biological information in listing determinations requires consideration of the scientific rigor of the methodologies utilized to develop any such information~~,~~ and the strength of that information. Rigor refers to the demonstrated validity of sampling, analytical, and assessment protocols and the availability of meta data in support of those protocols. Strength refers to the quantity of data and the extent to which such data demonstrates clear and convincing evidence of attainment or non-attainment of standards.

Availability of physical or biological data indicating use impairment may also be used to support listing when chemical data is otherwise insufficient in and of itself. Greater weight is given to data that provides direct, quantifiable documentation of impairment as opposed to data developed using surrogate indicators or parameters.

1. Impairment Where the Cause is Unknown

The federal CWA defines pollution as "the man made or man induced alteration of the chemical, physical, biological and radiological integrity of water," CWA §502(19). Pollution may result from the introduction of pollutants or from causative factors other than pollutants. Pollutants, as defined in the CWA at §502(6) include "dredged spoil, solid waste, incinerator residue, sewage, garbage, sewage sludge, munitions, chemical wastes, biological materials, radioactive materials (except those regulated under Atomic Energy Act of 1954, as amended), heat, wrecked or discarded equipment, rock, sand, cellar dirt, and industrial, municipal, and agricultural waste discharged into water." Notwithstanding the federal definition cited above, certain radiological constituents are also regulated under the state's Water Quality Control Act and are considered to be pollutants.

TMDL development is required in those instances where one or more pollutants are the cause of non-attainment. TMDLs are not required where the impairment is the result of pollution that is not a pollutant.

Waterbodies with classified uses that are impaired but where it remains unclear whether the cause of impairment is attributable to pollutants as opposed to pollution will be provisionally listed on the Section 303(d) List. The list will include a notation identifying waterbodies that are provisionally listed.

For waterbodies that are provisionally listed on the Section 303(d) List, the division, in cooperation with other interested persons, will undertake water quality monitoring and/or other water quality studies and assessments to determine whether the cause of the impairment is a pollutant. There will be a general goal of making this determination within ten years of provisionally listing any waterbody. No TMDL will be developed for a provisionally listed waterbody unless and until it is determined that the cause of the impairment is one or more pollutants.

<u>Provisional Listing Process</u>
Once a segment is provisionally listed, the process to determine a cause will include the following steps:

a. Determine if the impairment is caused by a pollutant.

Once a segment is provisionally listed on the 303(d) List, the cause of impairment must be identified through additional data collection and investigation.

If the impairment is caused by an identified pollutant, the segment will be placed on the 303(d) List as impaired for that pollutant (e.g., Cd, Fe(dis)) as well as for the applicable classified use without the provisional label. The division will proceed with development of a TMDL.

b. Determine if the impairment is caused by pollution.

If the evaluation demonstrates that the segment impairment is due to pollution, then the segment will be removed from the 303(d) List (as provisionally listed) and placed in the Integrated Report Category 4c (impairment is not caused by a pollutant) at the time of the next Section 303(d) List review cycle.

c. The cause of the segment impairment remains unknown.

If it cannot be determined that the cause is one or more pollutants or is not caused by pollution the segment will remain on the 303(d) List as provisionally impaired (e.g., aquatic life use, provisional). The cause of the impairment is to be determined within the next ten years.

A waterbody that is provisionally listed will not result in the prohibition of new or expanded discharges into the segment prior to the determination whether the impairment is caused by a pollutant.

To the extent it is suspected that a pollutant is the cause of the impairment, but the identity of the specific pollutant is not yet known, the waterbody segment will be provisionally listed. However, the fact that the waterbody is listed will not result in a prohibition of new or expanded discharges into the segment until the pollutant is identified.

2. Impairment Where the Source of the Pollutant is Natural

In cases where adequate monitoring and assessment indicate that natural conditions are the key factor of criteria exceedance(s), 303(d) listings will still be determined for impaired waterbody segments, as appropriate, without consideration of natural vs. anthropogenic causes. If natural conditions are triggering the exceedance(s), the decision is made by the commission through regulatory changes to the basin regulations in the triennial review process. Changes could involve the development of site specific standards or use removal through a use attainability analysis (UAA). The one exception in which anthropogenic causes are considered in determination of impairment is temperature assessments. The shoulder season excursion only applies if the temperature exceedances are not a result of anthropogenic sources (see below - III.D.3.e).

3. Impairment of Numeric Standards – Streams

Attainment of numeric chemical standards is assessed by comparison of ambient water quality against assigned standards. Assessment of chemical data considers attainment of both chronic and acute aquatic life use based chemical standards, where both chronic and acute standards have been assigned to a given waterbody.

a. Chronic Standards

Chronic standards can be evaluated one of two ways:

1. A comparison of the assessment statistic for that parameter (50th, 85th percentile – see III.C.1.a) to the standard, using the average hardness.
2. An evaluation of individual paired hardness and concentration assessments. The segment is considered not attaining if the paired calculations exceed more frequently than fifteen percent of the time.

In the case where both assessment approaches are conducted and the listing decision differs between the two, the paired hardness/concentration assessment decision is considered more representative.

b. Acute Standards

Acute standards are assessed by comparison of individual sample values against the standard. For the assessment of many metals standards, paired hardness and concentration values are needed. In general, data indicates non-attainment of an acute standard if the standard is exceeded more frequently than once in three years.

c. Agriculture and Domestic Water Supply Use Based Standards

These standards are expressed in terms of either one day or 30 day averaging periods (comparable to acute and chronic based standards, respectively) and are assessed by comparison of the percentile ranges (described in Section III.C.1.a. Chemical Data – General) against the standard. For assessments of standards listed as total fractions but where total species data is not available, the dissolved metals fraction is used to evaluate the standard as a conservative approach. To evaluate total standards expressed as 30 day averages, the 50th percentile of the dissolved data is assessed. For evaluation of standards expressed as one day averages, the individual dissolved values are compared to the standard. For assessment of standards expressed as one day averages, data indicates nonattainment if the standard is exceeded more frequently than once in three years.

d. Water Supply Use Based Standards

Nitrate/nitrite and arsenic: Assessment of nitrate and/or nitrite, and arsenic water supply use based standards will consider the combined total or individual ambient concentrations for nitrate and/or nitrite and the individual ambient concentrations for arsenic. Nitrate, nitrite and arsenic

standards are assessed along the entire segment for those segments where a water supply use has been adopted, regardless of whether or not there is a point of intake identified on the stream. The assessment will consider assessments and data used in permits development and will portion the impaired segment accordingly.

Where a range is specified for the arsenic standard, waterbodies will be considered in attainment of this standard, and not included on the Section 303(d) List, so long as the existing ambient quality does not exceed the second number in the range.

Manganese, iron and sulfate: For segments with adopted iron, manganese and sulfate water supply standards, the less restrictive of the following two options shall be applied as the numeric standard for assessment for the 303(d) List:

i. Table Value Standard - Iron (Dissolved) (Fe-D) – 300 ug/L
Table Value Standard - Manganese (Dissolved) (Mn-D) - 50 ug/L
Table Value Standard - Sulfate (SO_4) – 250 mg/L

Or

ii. Existing quality as of January 1, 2000.

To determine the existing quality as of the year 2000, a segment-specific library of water quality data has been created for these parameters. This 2000 water supply data library contains data from January 1, 1995 to December 31, 2009 and has been compiled from the following data sources: division, United States Geologic Survey (USGS) and River Watch Colorado (RW).

The division will aggregate data from 1995-1999 to characterize conditions from the year 2000. If less than ten samples are available from this time period, the period is extended using 5 year increments until at least ten data points are available. Note, data from the data library can only be considered representative of conditions from January 2000 if no new or increased sources of iron, manganese, and/or sulfate are known in the segment. If changes related to iron, manganese, and/or sulfate are known, only data collected prior to those changes can be used in assessments.

Determining the least restrictive standards for dissolved iron, dissolved manganese and sulfate and assessing current water quality requires the following steps:

1. As a screen, compare water quality data (the 50th percentile for sulfate or the 85th percentile for dissolved iron and dissolved manganese) from the current assessment period to table value standards (TVS).

2. For each parameter, if table value standards are not exceeded, the segment is considered attaining water quality standards. This concludes the assessment process for these parameters. Examination of water quality as of 2000 is not necessary in these cases.

3. For each parameter, if table value standards are exceeded, data for the segment in which the sampling station is located is compiled from the 2000 data library.

4. For segments with at least ten samples in the 2000 data library, the 85th percentile for each parameter is calculated to represent water quality as of 2000. This value is then compared to the table value standard and the least stringent value is used to evaluate attainment.

5. If no data representative of the year 2000 is available, or less than ten samples representative of the year 2000 are available for that segment, it is presumed that current ambient data is representative of existing quality as of January 1, 2000 and therefore the segment is not considered impaired (when supported by information that demonstrates there are no new or increased sources of iron, manganese, or sulfate). When not supported by information that demonstrates there are no new or increased sources of iron, manganese or sulfate, the segment will be placed on the M&E List due to uncertainty in standards/conditions from 2000.

6. If the table value standard is greater than the water quality as of 2000 value, and the data exceeds the table value standard, the segment is considered impaired. This concludes the assessment of these parameters.

7. If the water quality as of 2000 value is greater than the parameter's table value standard, a face value approach in which existing quality (50th percentile for sulfate or the 85th percentile for dissolved iron and dissolved manganese) from the current period of record is compared to the 50th or the 85th percentile of water quality data representing conditions as of the year 2000. If existing quality from the period of record exceeds conditions from 2000, the segment is considered impaired.

e. Temperature Standards
Temperature standards are assessed using the following three approaches:

i. Chronic temperature standard (MWAT): In general, data indicates non-attainment of a chronic temperature standard if the standard is exceeded by a weekly average temperature (as defined in Section C.5.a.i) more frequently than once in three years.

ii. Acute temperature standard (DM): In general, data indicates non-attainment of an acute temperature standard if the standard is exceeded by a daily maximum (as defined in Section C.5.a.ii) more frequently than once in three years.

iii. Excursions from temperature standard: In Regulation No. 31, Footnote 5c, Table 1 includes 4 excursions when exceedances of the temperature standard are acceptable. The party proposing the temperature listing (the proponent) is responsible for investigating temperature excursions. The following bullets outline the data necessary to demonstrate that these excursions apply:

- Air temperature excursion: Ambient water temperature may exceed the temperature criteria or the applicable site specific standard when the daily maximum air temperature exceeds the 90th percentile value of the daily maximum air temperatures in a given month calculated using at least 10 years of air temperature data. Nearby representative air monitoring stations will be used to assess this excursion.

- Low flow excursion: Ambient water temperature may exceed the temperature criteria or the applicable site specific standard when the daily stream flow falls below the acute critical low flow or monthly average stream flow falls below the chronic critical low flow, calculated pursuant to Regulation #31.9(1). Nearby representative gaging stations will be used to evaluate low flow conditions.

- Lakes and reservoirs excursion: Ambient water temperature may exceed the temperature criteria or the applicable site specific standard in the upper portion if there is adequate refuge. This excursion is discussed below in Section III.D.4.a.

- Winter shoulder season excursion: For the purposes of assessment, ambient water temperatures in cold streams may exceed applicable winter standards for 30 days before the winter/summer transition, and 30 days after the summer/winter transition, provided that the natural seasonal progression of temperature is maintained and that temperature exceedances during these periods are not the result of anthropogenic activities in the watershed. Examples of anthropogenic contributions include wastewater discharges, cooling water and runoff from impervious surfaces. The shoulder season excursion is applied to site-specific temperature tiers in the same way it applies to table value standards. A 30 day shoulder season is assessed for each tier.

If the temperature standard for a segment has been exceeded but the excursion data has not yet been evaluated, the segment will be placed on the M&E List.

4. Impairment of Numeric Standards – Lakes and Reservoirs

a. Temperature
For lakes and reservoirs, temperature measurements collected along a profile are assumed to represent the maximum weekly average temperature (MWAT). When a lake or reservoir is stratified, the upper portion may exceed the applicable standards, provided that adequate refuge exists in water below the upper portion. Adequate refuge depends on concurrent attainment within a given profile of the temperature standard and applicable dissolved oxygen standard. If the refuge is not adequate because of low DO levels, the lake or reservoir will be listed as impaired for DO rather than for temperature.

b. Dissolved Oxygen
If the average DO concentration in the upper portion of a lake falls below the standard in any profile, the lake will be listed. DO may be examined below the upper portion of the lake to determine if adequate refuge is available when assessing the temperature standard, as defined above.

c. pH
Assessment of pH data follows the general approach outlined above for stream samples, except that upper and lower portions should be assessed separately. First, the average pH is calculated from the upper and lower portions (as defined above) separately for each day. Then, the 15th and 85th percentiles of the average pH values for the period of record are compared to the minimum and maximum pH standard for both the upper and lower portions. Failure to attain the standard in either layer results in 303(d) Listing.

d. Metals and Inorganics
For these constituents, the assessment in lakes follows the same guidelines applied to stream samples except that samples from the upper and lower portions should be considered separately. Failure to attain the standard in either layer results in 303(d) Listing.

e. Site Specific Standards in Control Regulations
Some lakes and reservoirs have been assigned site specific standards for nutrients (total phosphorus) and chlorophyll a. These site specific standards are identified in control regulations which are specific to a given waterbody. These presently include Dillon Reservoir, Cherry Creek Reservoir, Chatfield Reservoir, and Bear Creek Reservoir, which are evaluated on an annual basis for compliance with site specific standards. Usually, the period for application of site specific standards is defined as the growing season and is described in the statement of basis and purpose for that standard. For example, growing season data are used to determine attainment with standards for phosphorus. Any determination of site specific standards attainment must be based upon application of such standards in a manner consistent with the applicable control regulation.

f. Multiple Profiles
If multiple profiles are collected from various locations for a lake on the same day, each profile will be evaluated separately. Multiple profiles will not be averaged for assessment purposes. If the division determines impairment is

isolated to an appropriate sub-segment or portion of a lake, the division may place the portion on the 303(d) List.

5. Impairment of Numeric Standards – General

a. Representative Data

Factors to consider when determining whether or not data is representative include: spatial distribution of sampling locations within the waterbody/segment, temporal variability of the data, changes in the watershed (i.e. changes in predominant land use, presence of new discharges, source removal or remediation projects), age of the data, method detection limits, bias in sampling design, etc.

During the assessment of sediment and MMI/aquatic life data, there may be situations where the most recent score is failing, however the vast majority of previous scores are attaining. In these cases, extra consideration will be used to examine the representative nature of the data. If the quality or representativeness of the data is in question, the segment will be proposed for the M&E List so that additional data can be collected.

For lakes and reservoirs, if a single profile indicates impairment but numerous attaining profiles exist in the dataset and other indications of impairment are absent, the division may place the lake on the M&E List so that additional data can be collected. If less than three profiles are available for assessment, the division will use best professional judgment to determine if the data is representative. If the quality of the data is in question, the lake or reservoir will be placed on the M&E List so that additional data can be collected.

b. Portions of Segments

If evaluation of a data set for an entire segment does not indicate impairment, but specific location(s) within the segment consistently exceed acute or chronic standards, the specific portion of the segment may be listed. This may also apply to lakes and reservoirs where sufficient data indicates impaired conditions are isolated to a specific portion of the lake. Segment portioning may also apply to those streams with MMI scores which demonstrate impairment, but not for the entire segment. Portioning for aquatic life using MMI scores will be decided on a case by case basis following Section III.D.7.a. and must include representative samples within the same sampling index period.

c. Evaluation of *E. coli* Standards

In 2010, the commission adopted a two month averaging period for the existing *E. coli* criteria. Evaluation of the *E. coli* standard is over multiple fixed two month intervals. The evaluation intervals are: January/February, March/April, May/June, July/August, September/October and November/December. A sample size of five or more is required for assessment of the two month intervals. Data will be assessed for each year if adequate data from each two month interval for any given year are available. If adequate data are not available to make an attainment decision using yearly data, then the division will assess *E. coli* data for that two month interval over the entire period of record.

If adequate data from two month intervals are not available to make an attainment decision, then assessment of the data will be on a seasonal basis. Because recreation typically occurs in the summer, the season of May through October will be used unless there is evidence that a different season is more appropriate.

Data sets comprised of two to four samples that indicate impairment of the *E. coli* standard will result in placement on the M&E List. Segments with *E. coli* data sets comprised of five to ten samples where there is overwhelming evidence of non-attainment (see section III.D.5.f - Overwhelming Evidence) will be placed on the 303(d) List. Data sets of more than ten samples indicating any degree of non-attainment, will also result in inclusion on the 303(d) List.

d. Temporary Modifications

When temporary modifications of numeric standards have been adopted, attainment is assessed against the underlying standard, including those instances where the decision to assign a temporary modification is based specifically upon significant uncertainty as to the appropriate underlying standard (see section 31.7(3)(a)(ii)(A) of the Basic Standards).

e. Sample size

Data sets comprised of two or three samples that indicate impairment of the chronic standard will result in placement on the M&E List except as noted for ambient-based standards, lakes and reservoirs below. Data sets comprised of four to nine samples where there is overwhelming evidence of non-attainment, or data sets of greater than or equal to ten samples indicating any degree of non-attainment, will result in inclusion on the 303(d) List except as noted for ambient-based standards, lakes and reservoirs.

For lakes and reservoirs, a minimum of five samples from the same location on different dates are required for the assessment of the metals and inorganic standards. If two to four samples indicate impairment, the lake is placed on the M&E List so that additional data can be collected. If the sample size is three or four and there is overwhelming evidence of impairment (see below), the waterbody will be placed on the 303(d) List. For the assessment of lake temperature and dissolved oxygen data, only a single profile is required for assessment.

For the assessment of ambient-based standards, a minimum of 10 samples is needed for conclusion of impairment (see Appendix B), except for data sets comprised of five to nine samples where there is overwhelming evidence of non-attainment will also result in a conclusion of impairment. In certain cases, where the sample size is less than five and the assessor suspects impairment, the division may place the segment on the M&E List.

f. Overwhelming Evidence

Overwhelming evidence consists of sufficient and credible data that clearly demonstrate that a waterbody's designated uses are impaired. Overwhelming evidence is demonstrated when representative data (data that accounts for temporal and spatial variation) indicates an exceedance of numeric water quality standards by more than 50 percent in magnitude.

g. Ambient Based Standards

Ambient based water quality standards are adopted where the table value standard cannot be met as a result of either natural conditions or irreversible, man induced conditions. Each ambient based standard is a site-specific characterization of existing quality derived from available representative data. To assess attainment of ambient based standards, the division uses a statistical approach based on the concept of the confidence interval to minimize uncertainty of assessment conclusions. If the lower confidence limit of the assessed value (e.g., 85th percentile) exceeds the standard, then the assessed concentration is significantly larger than the standard, and there is a high degree of confidence (95%) that the segment should be considered impaired. Without this statistical approach, applicable only to ambient based standards, there would be a much greater risk of incorrectly reaching an impairment decision. Because it has already been established that TVS cannot be attained due to natural or irreversible, man induced conditions, the evidence for further degradation (sufficient to warrant investment in a TMDL) should be especially compelling. Appendix B, *Assessing Attainment of Ambient Based Water Quality Standards in Colorado* includes a detailed description of the statistical basis of this approach with examples of the assessment procedures and tables used to determine the lower confidence limit. The appendix also includes a description of other legitimate assessment methods, which may be considered on a case-by-case basis if it can be shown that the alternative approaches are more applicable or appropriate for the dataset in question.

6. Impairment of Numeric Nutrient Standards

Prior to May 31, 2022, interim nutrient values will be considered for adoption by the commission only in the limited circumstances defined in Regulation No. 31 section 31.17(e). These include headwaters, direct use water supply (DUWS) lakes and reservoirs and other special circumstances as determined by the commission. In these limited cases, the commission may decide to adopt numeric nutrient standards for total phosphorus and chlorophyll *a* in a phased approach as basins are reviewed within the triennial review process. After May 31, 2017, total nitrogen will also be considered for adoption per Section 31.17(e).

a. Lakes

Nutrient concentrations in lakes are assessed as the seasonal average of values from the mixed layer, subject to provisions in Section III.C.2. When samples are collected from multiple depths in the mixed layer on the same date, the median of those values will represent the assessed concentration for that date. The annual seasonal average concentrations are compared against the standards with an allowable exceedence frequency of once every five years. In instances where the average nutrient concentrations exceed the standard but where there are fewer than three representative samples in a given season, those lakes will be placed on the M&E List until additional data can be collected. For lakes and reservoirs designated as DUWS, a minimum of five representative samples in a season are required for the assessments of chlorophyll *a* for that year.

Additionally, there are seasonal boundaries for the data used in the assessment process. For lakes and reservoirs designated as DUWS, chlorophyll *a* must be

collected from March 1 through November 20. For all other lake standards, samples should be collected during the summer season July 1 through September 30. Data collected outside of these times are not to be considered in the assessment for the 303(d) List.

b. Rivers and Streams

For rivers and streams, the assessed total phosphorus and total nitrogen concentrations are the annual median values. When compared to the nutrient standards, these values have an allowable exceedence frequency of once in any five year period. In cases where the average nutrient concentrations exceed the standard but where there are fewer than five samples from that year, those streams will be placed on the M&E List until additional data can be collected.

For rivers and streams, chlorophyll *a* (a measure of the areal abundance of attached algae or periphyton) is assessed during the summer season from July 1 to September 30. Assessment is based on the summertime maximum, "consistent with its foundation in a study of public responses to snapshot observations" (31.50(III)(B)). The allowable exceedence frequency is once in five years. Unlike the assessment of stream nutrient concentrations, only one sample is required for the assessment of stream chlorophyll *a*. However, the attainment of chlorophyll *a* for streams can be assessed where a representative sample can be obtained with the division's sampling protocol, which is designed for hard substrate. This protocol is titled *Standard Operating Procedures for the Collection of Periphyton Samples*, and is available upon request).

7. Impairment of Narrative Standards and Classified Uses

Impairment of narrative standards and classified uses may be supported by chemical data and/or information generated by biological and/or physical assessments. In instances where a determination of impairment is based solely upon biological and/or physical assessments, such assessments must provide clear and convincing evidence of non-attainment.

a. Aquatic Life Use

For aquatic life uses, the division will generally consider impairment of narrative standards or classified uses to be demonstrated when either the physical/habitat data or biological community metrics reflect a condition that is significantly less than the expected or reference condition. The division will also consider an impairment of the aquatic life use if a showing has been made consistent with the protocols established in the commission Policy 10-1, *Aquatic Life Use Attainment, Methodology to Determine Use Attainment for Rivers and Streams.* The commission Policy 10-1 is available either hard copy or electronically on the Colorado Water Quality Control Commission web site (https://www.colorado.gov/pacific/cdphe/wqcc-policies).

In regards to the aquatic life sampling season index period, the division recommends following the sampling protocols outlined in Appendix B of Policy 10-1. Appendix B outlines a recommended standard index period of July 1 to October 1, and May 1 to October 1 for elevations less than 1,550 meters. However, the division recognizes that uncharacteristic flow conditions, including

but not limited to, seasonal irrigation releases and return flows, often preclude sample representativeness and sampler safety within the recommended index periods. In these circumstances, the division will consider samples collected up to four weeks after the recommended sampling season index period on a case by case basis.

Benthic data submitted to the division to be used in 303(d) listing decisions using Policy 10-1 is to be submitted by the data cutoff date and in the preferred data format as requested by the division. Upon receipt of all data, the division will update the benthics master taxa table and will send it to the organization for concurrence on taxa. The division will then conduct a single sub-sampling[1] of each sample. Once the subsampling is completed by the division that version of EDAS (Ecological Data Application System) can be sent to all parties. MMI scores and metrics will be generated using this version of EDAS.

The division will consider all reliable and representative aquatic life data including information regarding other assemblages (e.g., fish or algae) in determining whether or not a stream is impaired. In these cases, the criteria specified in Section III.C.3, above, will be used.

For those segments with differing MMI scores (one passing MMI score; one failing MMI score) on the same representative portion of a segment or station, taken within the same calendar year sampling season, this segment would be included on the M&E List. For those segments with differing MMI scores on the same representative segment taken in different calendar year sampling seasons, the more recent MMI score will be the one used in the listing decision. The division will consider the representative nature of all aquatic life data before listing decisions are made. Clear and convincing evidence is required to show impairment.

In cases where MMI scores for multiple biotypes are considered in the assessment of the segment, impairment is determined when both MMIs have failing scores. If one MMI passes and the other fails, the segment can be concluded as impaired based on additional supporting information and justification. Otherwise, the sample data for the site is deemed inconclusive with respect to determining impairment.

Upon determination of impairment, the waterbody will be included on the 303(d) List unless the segment is currently included in Integrated Reporting Category 4a, 4b or 4c for an aquatic life use standards impairment. If the segment is previously listed, or proposed to be listed for a pollutant causing an impairment of the aquatic life use, the segment will be listed for that pollutant as well as for impairment of the aquatic life use.

If there is no apparent pollutant, the impairment will be identified as provisionally listed.

---

1 A subsample as defined in the context of the Colorado MMI, is a procedure to reduce the individual organism count of a whole sample exceeding 360 individuals, to a standardized, 300-fixed organism count between 240 and 360 individuals.

Consistent with listing methodology protocols, more data is not necessary to remove a segment than the amount of data used to list the segment. If one MMI score was used to list a segment, then a single, more recent, reliable and representative MMI score is sufficient to remove the segment from the 303(d) list for attainment. For high scoring waters (>64 for biotypes 1 & 2; >44 for biotype 3), a 22 point decline in the MMI score results in a conclusion of impairment. These segments will then be determined to be in attainment when subsequent MMI scores improve by a minimum of half of the original decline and exceed the attainment threshold (as a station to station comparison).

b. Water Supply Use

For water supply uses, the division will consider chemical data, biological and/or physical assessments that provide clear and convincing evidence of non-attainment. Such impairment may be demonstrated by chemical data documented at levels toxic to humans. The division will utilize commission Policy 96-2, *Human Health Based Water Quality Criteria and Standards*, in any determination of impairment based upon such information. Impairment decisions may also be supported by biological and physical data presenting overwhelming evidence of impairment due to color, taste and odor.

c. Narrative Free from Toxics Standard

In-situ bioassay, or other ambient toxicity test results which demonstrate statistically significant lethal or sub-lethal adverse effects and which are supported by biological information demonstrating adverse impacts to aquatic community health, composition, or productivity, in comparison to an appropriate reference condition, will result in a decision of impairment. In general, interpretation of toxicity test results will conform to applicable portions of the *Implementation of the Narrative Standard for Toxicity In Discharge Permits Using Whole Effluent Toxicity (Wet) Testing* (https://www.colorado.gov/pacific/cdphe/water-quality-permitting-policies).

For lakes and reservoirs, impairment may be demonstrated where acute conditions (typically low DO levels) result in significant fish kills. Fish kills associated with accidental spills or isolated unauthorized discharges of toxics will not typically be considered a basis for listing.

d. Narrative Sediment Standards

Excessive deposition of sediment on the bottom of streams and rivers can cause harmful impacts to aquatic life such as benthic macroinvertebrates and fish, in addition to other beneficial uses. The impacts to aquatic life usually result from the loss of critical habitat for fish, aquatic invertebrates and algae. Regulation #31 includes a narrative standard that states that a waterbody should be "free from substances attributable to human-caused point source or nonpoint source discharge in amounts, concentrations, or combinations which can settle to form bottom deposits detrimental to the beneficial uses."

The division determines attainment of the statewide narrative standard by following protocols outlined in commission Policy 98-1, *Guidance for Implementation of Colorado's Narrative Sediment Standard Regulation #31, Section 31.11(1)(a)(i)*. For all state waters, the narrative standard is not in attainment when evidence demonstrates the following:

- The actual observed sedimentation condition for a specific waterbody is significantly different than the expected condition, and thus considered excess sediment
- The excess sediment is attributable to an anthropogenic source
- The excess sediment could be a detriment to a beneficial use

Policy 98-1 includes sediment thresholds that apply to rivers and streams in specific regions, as well as specific assessment methods to evaluate i) benthic macroinvertebrate assemblages and ii) fish assemblages.

i. To evaluate the benthic macroinvertebrate assemblages, three components are examined:

   - a census of the waterbody substrate and a resultant measure of the percent fines (%fines <2 mm),
   - a Tolerance Indicator Value for sediment (TIV_{SED} score), and
   - a review of available watershed information (watershed review).

   A detailed explanation of how each component is evaluated is included in Policy 98-1. Sediment and macroinvertebrate data used to make attainment decisions must be collected within the same two week period during representative flow conditions. For a segment to be in non-attainment, a failing TIV_{SED} score, a failing % fines value and a watershed review are required. The watershed review must confirm the existence of anthropogenic sources of sediment and confirm that the sample site/watershed is not significantly different from the range of conditions used to establish the expected condition for the Sediment Region. Impairment decisions are not possible if only two of the three components are assessed. The TIV_{SED} score and the % fines must be in attainment in order for the division to propose a delisting of a previously listed segment.

ii. To evaluate fish assemblages, the percent fines (percent fines <8 mm) is measured from targeted fish spawning habitat for a given segment. If the percent fines is greater than 20 percent and the watershed review confirms that excess sediment is attributable to an anthropogenic source, the segment is considered impaired.

The division will only consider reliable and representative data in determining whether a segment is impaired. Data collected from a single representative location and time period is sufficient to make an attainment decision. Consequently, data from a single more recent representative sampling event can also result in a segment de-listing. If multiple data sets are collected from the same location within a five year period of record, the most recent representative data set is used for attainment decisions.

The extent of the impairment is determined as a part of the watershed review in the assessment process. The division considers watershed characteristics in determining whether to list only a portion of a segment consistent with methodology described in this guidance in section III.D.6.b.

Determinations regarding impairment of beneficial uses other than aquatic life will be made in accordance with Section V of Policy 98-1.

8. Listing Based on Elevated Mercury in Fish Tissue

Waterbodies are assessed for attainment of Colorado's aquatic life use (EPA's fishable goals of CWA Section 101(a)(2)) by comparing the weighted average fish tissue mercury for each species/size class to a 0.3 parts per million (ppm) threshold level. For small datasets with a large portion of the data below the detection limit, the division will substitute half the detection limit when calculating the weighted average. Those waterbodies with a weighted average fish mercury concentration for each species/size class that exceeds the 0.3 ppm threshold level will be placed on the 303(d) List. A minimum of 30 fish tissue samples (either as individual samples or composites) from each species/size class are necessary to determine impairment of a waterbody for mercury in the fish tissue. For waterbodies where the data is short of this requirement, the waterbody will be placed on the M&E List so that additional data can be collected for assessment. If the sample size is between 10 and 30 and the weighted average fish tissue mercury concentration is greater than 1.5 times the threshold level, the waterbody will be placed on the 303(d) List based on overwhelming evidence of impairment.

Those waters that are listed due to elevated levels of mercury in fish tissue may be identified as low priority (notwithstanding the provisions of Section IV.B.1 below) when the provisions applicable to EPA reporting Category 5m are satisfied (see *Information Concerning 2008 Clean Water Act Sections 303(d), 305(b) and 314 Integrated Reporting and Listing Decisions,* EPA, October 12, 2006). Waters are placed in reporting Category 5m in instances where impairment is due to atmospheric deposition and where the state has a comprehensive mercury control program in place. The division will evaluate each listing arising from mercury levels in fish tissue for evidence of current and historic mining activities within the contributing watershed, for other potential industrial sources and for potential geologic influences.

IV. PRIORITIZATION FOR TMDL DEVELOPMENT

The division must ensure that TMDLs are developed for all waterbodies and pollutants on the Section 303(d) List. Recognizing that all TMDLs cannot be completed at once and that certain risks may be greater than others, the CWA directs the division to prioritize the waters on the Section 303(d) List. The division will use the prioritized Section 303(d) List to focus resources to establish priority waters or watersheds and support the development of targeted TMDLs. Provisionally listed segments will not be prioritized for TMDL development.

A. Prioritization Objective

The objective of the prioritization on Section 303(d) list is to identify where the division and the public should focus their resources. The identification of high priority segments do not necessarily mean that the TMDLs will be developed before any lower

priority segments. For some high priority segments, the development of a TMDL may be delayed due to the need for additional data collection or stakeholder outreach.

B. Assigning Priorities

Priorities defined on the 303(d) list are initially based on consideration of the severity of impairment to the use classifications for the segment. Use classifications are described in *Basic Standards and Methodologies for Surface Water,* Regulation No. 31 (5 CCR 1002-8, sec. 31.13). Secondary factors can be used to modify the initial prioritization to an overall or final prioritization. Secondary factors may either elevate a waterbody into a higher priority group e.g., endangered or declining native species, public interest, administrative needs or reduce the priority ranking e.g., pace of stakeholder group development, CERCLA cleanup action in progress.

1. Severity of Water Quality Impairment

   High Priority: Non-supporting for water supply standards based on Safe Drinking Water Act primary drinking water standards (NO_2, NO_3, As) Aquatic Life Class 1 cold or warm, or Recreation Use Class E. Listings based on high levels of mercury in fish tissue.
   Medium Priority: Non-supporting for Aquatic Life Class 2 cold or warm, or Agriculture.
   Low Priority: Non-supporting for other water supply standards or Recreation Use Class P, U or N, or non-supporting for underlying standard where a temporary modification based specifically upon significant uncertainty as to the appropriate underlying standard has been adopted and the commission has determined that there is an appropriate plan in place to resolve the uncertainty.

2. Secondary Considerations

   - Division action can support a local, regional or federal stakeholder group that is ready to move on to the next step of TMDL development or there is substantial public interest and support.
   - The waterbody is vulnerable or fragile as an aquatic habitat or there are aquatic species of special concern present.
   - The waterbody is of particular importance for recreational, economic and aesthetic uses.
   - The division can realize efficiency savings (e.g., synchronizing permits, linking segments within a watershed, availability of water quality data).
   - There are immediate programmatic needs such as waste load allocations for permits that are due to expire, ~~or~~ for new or expanding discharges, or to facilitate 319 project developments in priority watersheds.
   - There is a court ordered cleanup or CERCLA action in progress, which will change the contribution of pollutants (this consideration could reduce priority ranking).

3. Identification of Targeted TMDLs and Priority Waters or Watersheds

   It is the division's intent that TMDLs that are designated as targeted TMDLs will be completed prior to the next listing cycle, or within two years of finalizing

the 303(d) List by the commission. Targeted TMDLs will most likely be included in priority waters or watersheds that are designated through a prioritization framework using multiple factors including, but not limited to, the high priority waters for TMDL development as defined in the 303(d) List. However, not all high priority listings are suitable for TMDL development within a two year window. For example, adequate data to support TMDL development is not available for all high priority listings. Conversely, waters designated as medium or low priority may be amenable to TMDL development within the next two years and therefore may be targeted for TMDL development at this point.

TMDL development is subject to a variety of factors that are both within and beyond the division's control. These may include availability of adequate data, local or broader political concerns, new information that affects the listing decision, coordination with remedial programs such as CERCLA or Superfund, or availability of division resources. Designation of a TMDL as targeted should be considered for planning purposes, but should not be treated as a definitive division workplan commitment. The division's TMDL program workplan is updated quarterly and is available on the division website (https://www.colorado.gov/pacific/cdphe/total-maximum-daily-loads-tmdls).

V. POLICY DETERMINATIONS

A. 2016 Listing Methodology

In March 2015, the commission made the following policy decisions as they relate to assessment methods used to determine impaired waters in Colorado.

Category 4c: The commission determined that Category 4c could be a useful tool for identifying segments that are impaired solely due to pollution and not attributable to pollutants. Segments placed in Category 4c will generally not require TMDL development but may require pollution reductions plans to address the impairment. The commission directed the division to work with interested stakeholders to develop guidance for the determination of Category 4c segments as well as guidance for future plans for restoration on these segments. Work on these issues will be presented in the 2018 303(d) Listing Methodology proposal in March 2017.

Assessment of the iron, manganese and sulfate water supply standards: The standards for iron, manganese and sulfate in the Basic Standards, Regulation #31 (circa 2015) are assessed using the least stringent standard: table value standards or the existing quality as of the year 2000. The commission clarified that when no data is available for the year 2000, current water quality should be presumed to be representative of existing water quality as of 2000 if a watershed review concludes that no additional sources of manganese, iron or sulfate have been introduced since 2000. Where there have been no changes in source contributions the segment is not impaired. Where there have been changes in source contributions, segments will be placed on the M&E List and considered high priority for site-specific standards development.

The commission also decided that in instances where the water quality representative of the year 2000 was used as the standard for listing decisions, the assessment of current

conditions is conducted using the statistics used to evaluate the table value standards, for manganese and iron it is the 85th percentile and for sulfate it is the 50th percentile.

The commission established that when setting a standard for iron, manganese and sulfate using data representative of existing quality in 2000, 10 data points is typically needed to have certainty in the standard being used.

Assessment of the nitrate, nitrite and arsenic water supply standards:
The commission adopted language regarding the assessment of arsenic and nitrate/nitrite based on the current standards. The commission intends to revisit this in the event that the standards are modified in the future.

INDEX

APPENDIX A

Spawning Table

Spawning	Egg Incubation	Intra-gravel Sac Fry	Fry Emerge

OCT NOV DEC JAN FEB MAR APR MAY JUN

Successful spawning is defined as all four phases or conditions that allow eggs to be deposited, incubated, hatch and fry emergence.

Table* of approximate time and duration of spawning, and critical early development life stages for brown and rainbow trout in 11 physical habitat simulation study streams.

*From: *Determination of Population Limits For Critical Salmonid Habitats in Colorado Streams Using Physical Habitat Simulation System,* B. Nerhing and R. Anderson. Rivers Journal, January 1993, Vol 4 No. 1, page 1-19, Table 3.

Gravel River	Species	Adult Spawning	Egg Incubation	Egg Hatching	Fry Emergence
Arkansas	brown	10/15 – 11/15	10/15 – 4/1	3/1 – 5/1	4/1 – 6/1
Blue	brown	10/15 – 11/15	10/15 – 6/1	4/1 – 6/1	5/15 – 7/1
Cache La Poudre	brown	10/15 – 11/15	10/15 – 6/1	4/1 – 6/1	5/15 – 7/1
Cache La Poudre	**rainbow**	**4/15 - 5/30**	**4/15 - 7/15**	**6/15 - 7/15**	**7/1 - 8/1**
Colorado	brown	10/15 – 11/15	10/15 – 4/1	4/1 – 6/1	5/15 – 6/15
Colorado	**rainbow**	**4/15 - 4/30**	**4/15 - 6/15**	**6/1 - 7/1**	**6/15 - 7/15**
Frying Pan	brown	10/15 – 11/15	10/15 – 5/1	4/1 – 6/1	5/15 – 6/15
Frying Pan	**rainbow**	4/1 – 5/1	4/1 – 6/15	6/1 – 7/1	6/15 – 7/15
Gunnison	brown	10/15 – 11/15	10/15 – 4/1	3/15 – 5/15	5/1 – 6/15
Gunnison	**rainbow**	**4/1 - 5/1**	**4/1 - 6/15**	**6/1 - 7/1**	**6/15 - 7/15**
Rio Grande	brown	10/15 – 11/15	10/15 – 5/1	4/1 – 6/1	5/15 – 6/15
S Fk Rio Grande	brown	10/15 – 11/15	10/15 – 6/1	5/1 – 7/1	6/1 – 7/15
S Platte	brown	10/15 – 11/15	10/15 – 5/1	4/1 – 6/1	5/1 – 6/15
S Platte	**rainbow**	**4/1 - 5/15**	**4/1 - 6/1**	**6/1 - 7/1**	**6/15 - 7/15**
St Vrain	brown	10/15 – 11/15	10/15 – 5/1	4/1 – 6/1	5/15 – 7/1
Taylor	brown	10/15 – 11/15	10/15 – 5/1	4/1 – 6/1	5/15 – 7/1

APPENDIX B

Assessing Attainment of Ambient-Based Water Quality Standards in Colorado

Ambient-based water quality standards have been adopted in Colorado in limited circumstances where the table value standard cannot be met as a result of either natural conditions or irreversible, man-induced conditions. Each ambient-based standard is a site-specific characterization of *existing quality*[1] derived from "available representative data". Once an ambient-based standard has been adopted, attainment is assessed using recent, representative data.

The mechanics of setting the ambient-based standard and assessing its attainment are the same, but the characterizations are carried out with different, possibly overlapping, data sets. For dissolved metals, for example, the chronic standard is set equal to the 85th percentile of the available, representative concentration data, and the acute standard is set equal to the 95th percentile[2]. When existing quality is assessed, the 85th percentile of the available, representative concentration data in a subsequent data collection is compared to the chronic standard, and the 95th percentile is compared to the acute standard.

Assessment determines if water quality continues to meet the level of ambient quality originally characterized by the standard. In the current assessment methodology, the same quality is maintained (i.e., the standard is attained) if the assessed value does not exceed the standard. If the assessed value exceeds the standard, water quality is considered to be impaired. McBride (2005) calls this a "face value" test because it does not include consideration of sampling error.

Current assessment methodology for ambient-based standards has proven problematic. Successive assessments may yield opposite conclusions about the maintenance of existing quality. Changed assessment conclusions can have significant practical ramifications when they cause water bodies with ambient-based standards to move on and off of the 303(d) list, as has happened in a number of cases (Table 1).

Table 1. Historical changes to listings based on ambient-based standards for selected stream segments.

Assessed	COARLA01a[3]	COARLA01b	COARLA01c	COARMA04a	COSPBD01
1998	List: Se, SO4, Fe	List: Se, SO4, Fe	List: Se, SO4, Fe	List: Se	

[1] 31.5(20) "EXISTING QUALITY" means the 85th percentile of the data for total ammonia, nitrate, and the dissolved metals, the 50th percentile for total recoverable metals, the 15th percentile of such data for dissolved oxygen, the geometric mean of such data for E. coli, and the range between the 15th and 85th percentiles for pH. For temperature, for the purposes of implementing the chronic standard, "existing quality" means the maximum WAT in a three year period.

[2] Concentrations corresponding to the specific percentiles are estimated from the available data using the PERCENTILE function in EXCEL. In most cases, these concentrations are interpolated values.

[3] COARLA01 was split into segments 1a, 1b, and 1c in 2002.

Assessed	COARLA01a[3]	COARLA01b	COARLA01c	COARMA04a	COSPBD01
2002	Delist: Se, SO4, Fe	Delist: Se, SO4			
2004	List: Se	List: Se; Delist: Fe	List: Se		
2006	List: Fe				List: Se
2007				Set ambient std	Set ambient std
2008	List: SO4, Delist: Fe			Delist: Se	
2010					Set seasonal std
2012				List: Se	Delist one season

In retrospect, it should not be surprising that successive assessments of ambient-based standards could yield different conclusions even in the absence of any water quality change. Successive assessments based on the same percentile (e.g., 85th) are affected by normal variability in the available concentration data. Seasonal patterns, stochastic variation, and sampling or analytical error all contribute to that variability. Consequently, we might expect about half of the assessed values to be larger than the standard and half smaller.

When an assessment shows that an ambient-based standard is exceeded by even a small amount, the water body may be placed on the 303(d) list. Assuming no trend in ambient concentrations, it is equally likely in the next assessment cycle that the assessed value will fall below the standard. When the assessed value for a listed water body falls below the standard, the water body is removed from the 303(d) list. Thus, the examples in Table 1 of water bodies going on and off of the 303(d) list is consistent with statistical expectations for the current assessment methodology.

Having water bodies move on and off the 303(d) list creates two problems. The first problem is that it takes time and effort to develop or revise the 303(d) list. The second problem is that listing has practical ramifications for dischargers. Both problems can be addressed by adding an explicit level of confidence to assessments of ambient-based standards. The addition of a defined level of confidence would not affect the underlying definitions of existing quality or ambient-based standards, but would establish the reliability of conclusions drawn from assessments.

Statistical Approach

Increasing the reliability of conclusions drawn from assessments of ambient-based standards is based on the statistical concept of the *confidence interval*. The confidence interval is often viewed as the region around an estimate (i.e., the assessed concentration) within which the true concentration (i.e., the standard) is thought to be located[4]. If the confidence interval of

[4] McBride (2005; p. 58) explains why this simplistic view is not strictly correct. Nevertheless, it is useful for communicating the approach.

the assessed concentration (e.g., 85th percentile) does not include the standard, then the assessed concentration is significantly different from the standard.

The width of the confidence interval, and thus the range of concentrations it spans, is determined in part by the desired level of confidence. When the level of confidence is set to 95%, for example, it means there is only a 5% probability (a 1-in-20 chance) of mistakenly concluding that the assessed concentration differs from the standard (i.e., a Type 1 error). Setting the risk of a mistake to 5% (a 1-in-20 chance) would improve the reliability of future assessments compared to the current approach.

A level of confidence other than 95% could be used, but there are tradeoffs. A higher level of confidence, such as 99%, has the advantage of reducing the risk of Type 1 errors to 1% instead of 5%, but it would also result in broader confidence intervals. Having broader confidence intervals makes it less likely that an exceedance will be identified because it is more likely that the standard will fall within the interval. Conversely, a lower level of confidence, such as 90%, yields a narrower confidence interval, but an increased risk (10% probability of a Type 1 error, instead of a 5% probability) of claiming that a segment is impaired when it is not.

For most assessments, regulators are interested only in situations where the assessed concentration is significantly larger than the standard. Thus, the null hypothesis can be defined to assume that the assessed concentration is less than or equal to the standard (i.e., H_0: Assessed concentration $\leq$ Standard); the test is one-sided. Rejection of the null hypothesis for this one-sided test means that the assessed concentration is significantly larger than the standard. In this one-sided case, a 5% probability of a Type 1 error defines the risk of claiming that a water body is impaired when it is not.

Selection of a specific statistical approach is affected by the number of assessments that the Division must undertake on a regular basis. If assessments were required only occasionally for one constituent at one site, there are a number of parametric and non-parametric tests to choose from. However, the Division must contend with about 170 ambient-based standards that have been adopted in 100 water bodies across the state. Running tests separately for each constituent in each water body during each assessment cycle would not be practical for the Division, but may be appropriate and acceptable for parties that may have a narrower focus.

One way to keep the large workload manageable is to define the confidence interval for the assessed value in terms of percentiles rather than concentrations. Defining the confidence interval of a percentile is inherently non-parametric and well-suited to typical assessment data sets for which the distribution usually is not known in advance. Furthermore, it would be difficult, especially with small sample sizes, to validate a distribution. If a distribution cannot be assumed or validated, application of parametric methods becomes questionable.

Confidence intervals for percentiles are a function only of sample size. Tables can be developed to define confidence intervals that would be applicable to any constituent at any site for which the assessed data set consisted of the same number of measured concentrations. Thus, an assessment for zinc in one watershed and one for copper in another

watershed would use the same confidence interval for percentiles as long as both sites had the same number of observations.

Assessment with percentiles is best understood with some graphical examples, and a good place to begin is with the current assessment methodology. The 85th percentile is featured in this example because it is the most common among ambient-based standards, but the concept applies equally well to the other percentiles (95th or 50th). The current methodology locates the concentration that corresponds to the 85th percentile[5] of the assessed data set and compares it to the standard. When the assessed value (85th percentile) is larger than the standard, current assessment methodology registers an exceedance (Figure 1). The magnitude of the exceedance, in terms of concentration, may be large or small, but the outcome is the same. However, the current methodology does not specify the reliability of the conclusion.

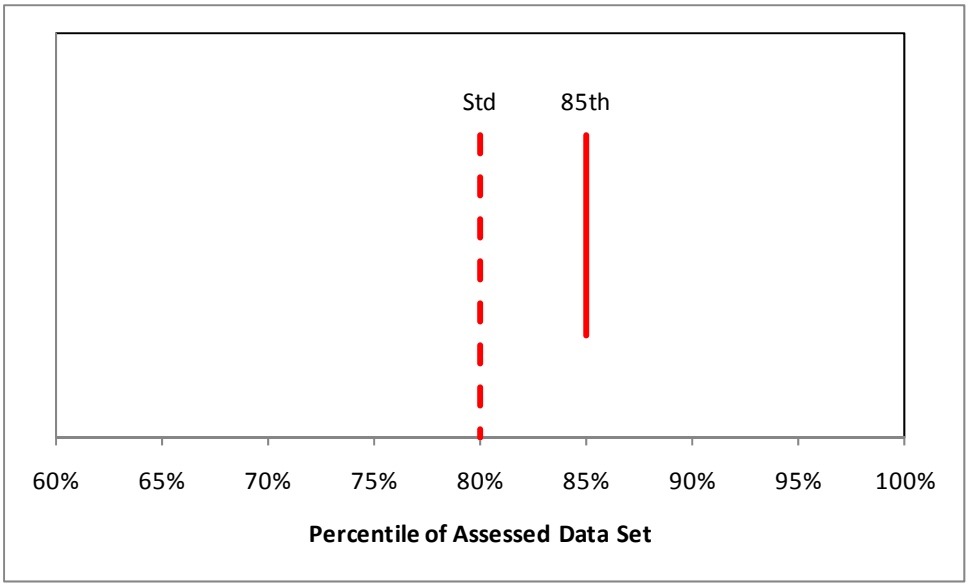


Figure 1. Graphical representation of a scenario where the 85th percentile of the assessed data (solid red line) exceeds the standard (dashed red line). For convenience in presenting the example, the concentration of the standard is represented as a percentile (80th) of the assessed data set. Current assessment methodology would interpret the result as an exceedance of the standard.

Reliability can be specified by defining a confidence interval around the 85th percentile. For example, when a confidence level of 95% is specified, the confidence interval constructed around the estimate (i.e., the 85th percentile) has a 95% probability of containing the true value (i.e., the standard). Making the right call 95% of the time is a very reliable basis for decision-making.

Building on the scenario used for Figure 1, a one-sided confidence interval is constructed for the 85th percentile of the assessed data. A statistical test is formalized with a null hypothesis

[5] The PERCENTILE function in EXCEL is used to determine the concentration that corresponds to the 85th percentile of all concentrations in the assessed data set. High concentrations correspond to high percentiles; thus, the 85th percentile is a high concentration within the assessed data set. For reasons to be explained later, the Excel function has shortcomings that should be considered in future assessments.

(H_0) stating that the 85th percentile of the assessed data set is less than or equal to the standard. The test is one-sided to determine if the assessed value (85th percentile) exceeds the standard because assessment is focused on exceedances. The null hypothesis is rejected when the LCL exceeds the standard.

In this example, the standard falls within the confidence interval for the 85th percentile (Figure 2). Thus, the concentration corresponding to the 85th percentile of the assessed data set is not larger than the standard. The null hypothesis is not rejected, and there would be no justification for listing the water body.

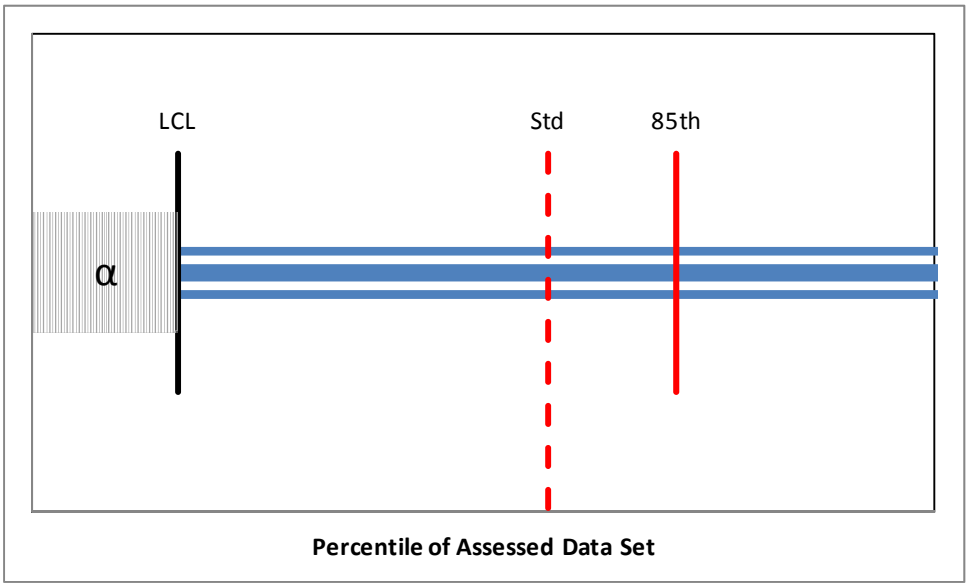


Figure 2. One-tailed confidence interval with the lower (LCL) confidence limit added to the scenario shown in Figure 1. Although the 85th percentile of the assessed data exceeds the standard, the difference is not statistically significant because the LCL of the 85th percentile does not exceed the standard. The critical region (gray region marked with alpha) extends to the left (lower percentiles) of the LCL. The confidence level for the interval is 1-α.

The scenario in Figure 2 is now changed so that the assessed value is significantly larger than the standard (Figure 3). The standard now corresponds to a low percentile of the assessed data distribution. With this scenario, the lower confidence limit (LCL) of the assessed value (85th percentile) exceeds the standard, and the null hypothesis is rejected. The assessed value is significantly greater than the standard, and the outcome would support a listing decision.

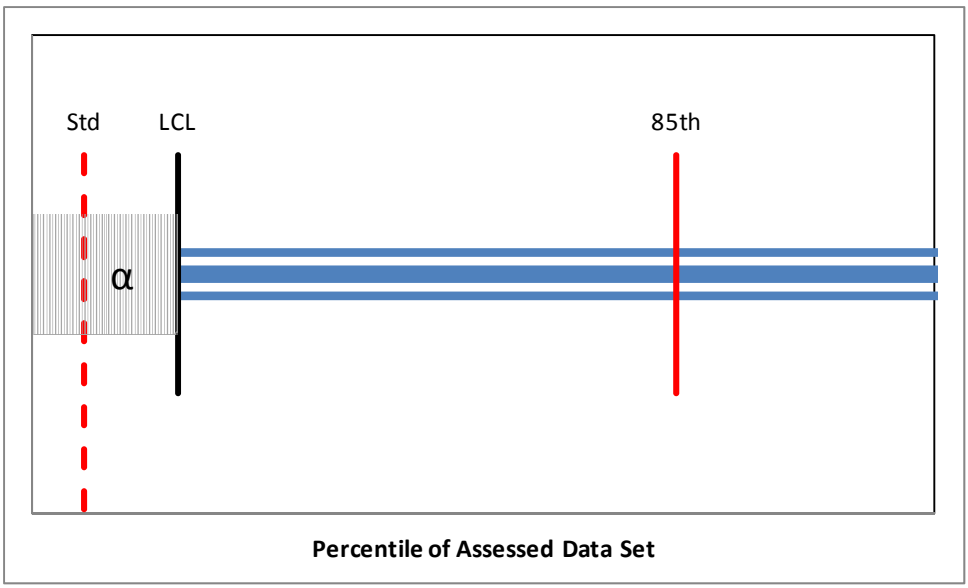


Figure 3. The scenario shown in Figure 2 is modified so that ambient concentrations are high relative to the standard. The LCL is greater than the standard, meaning that the 85th percentile of the assessed data is significantly greater than the standard.

Statistical Implementation

Defining the LCL is the key to a defensible statistical assessment of Colorado's ambient-based standards. There are a number of possible approaches (see Helsel and Hirsch 2002), but there are compelling reasons for the Division to focus on non-parametric methods, as explained below. A brief overview of common approaches provides a useful introduction to the concepts before tailoring an approach to our needs.

Overview of Methods

Variance and sample size are required for locating the LCL because the confidence interval is a statement about uncertainty. Since most environmental data sets are not normally distributed (or are too small to test for normality) a non-parametric test is preferred in most cases. A non-parametric method for locating the LCL makes no assumptions about the underlying statistical distribution of the data. Non-parametric methods for defining confidence intervals rely on the binomial distribution for defining variance. Exact and approximate non-parametric methods are available.

The Clopper-Pearson equations are used to determine exact confidence intervals for percentiles (Equation 1), but the computation is tedious, especially for large sample size. Less well-known, and much less tedious, is a direct calculation (Equation 2) using the F distribution (Leemis and Trivedi 1996), which can be evaluated in EXCEL with the FINV function.

APPENDIX B

$$\sum_{k=y}^{n} \binom{n}{k} p_L^k (1-p_L)^{n-k} = \alpha$$

Equation 1. Clopper-Pearson equation for the lower confidence interval. Probabilities are evaluated for successive proportions (p_L) until the sum reaches the desired exceedance level (α). Each proportion is calculated from the number of successes (k) out of the number of trials (n).

$$p_L = \frac{1}{1 + \frac{n-k+1}{kF_{2k,2(n-k+1),1-\alpha}}}$$

Equation 2. Exact lower confidence limit defined in terms of the F distribution; k, n, p_L, and α are defined in Equation 1. The result is identical to that obtained by iteration with Equation 1.

An exact solution for the confidence interval may sound like the ideal approach, but it is not well-aligned with Colorado's assessment needs, which would benefit from a target percentile that applies to data sets of any size. Because the binomial is a discrete distribution, assessments would logically be based on the [integer] number of samples that exceed the standard. However, the discreteness of the distribution precludes locating the LCL exactly for most sample sizes since the 85th percentile corresponds to an integer value of k only when sample size, n, is a multiple of 20[6]. When sample size is not a multiple of 20 and the confidence interval must be calculated with an integer, the resulting confidence interval would be larger than 95%. In other words, it becomes less likely that an exceedance will be identified.

Approximate methods also exist for defining confidence intervals for percentiles. In fact, many statisticians (e.g., Agresti and Coull 1998) recommend approximate methods because the exact method yields confidence intervals that tend to be too large (i.e., *exact* is something of a misnomer). A number of approximate methods have been developed for estimating confidence intervals. Historically, the Wald confidence interval has been recommended, especially when sample size is large[7]. It is also the easiest to understand.

The Wald test could be used to evaluate the null hypothesis that the estimated 85th percentile (i.e., the assessed value of a recent data set) is equal to the true 85th percentile (i.e., the standard, which characterizes existing quality). The difference between the estimate ($\hat{p}$) and the standard (p_0) is assumed to be approximately normally distributed at larger sample size. When the difference is divided by the standard error of the estimate (Equation 3), the result can be compared to standard normal deviates (z). Inverting the test yields a two-sided, 100(1-α)% confidence interval for p_0 (Equation 4).

[6] Given that the percentile is calculated as p=k/n, the first integer combination that delivers p=0.85 is when k=17 and n=20.

[7] A sample size of 30 or more is often regarded as large, but this rule of thumb may only be helpful where the central limit approximation is applicable. See Brown et al (2001) for a brief review of common sample size recommendations.

$$z = \frac{(\hat{p} - p_0)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Equation 3. Wald test statistic for the difference between estimated and true values of a percentile.

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Equation 4. Inversion of the Wald test statistic to yield a confidence interval for the estimated percentile.

For many years, the Wald interval was the recommended approach for large sample size. A perusal of the literature shows that there is not much agreement now on what constitutes "small", especially where percentiles are extreme (e.g., close to zero or to 100%). At small sample size, which is common for assessments, the Wald interval tends to be too small to accurately define the 95% confidence interval. However, concerns about the performance of the Wald interval extend beyond the issue of sample size. Brown et al (2001) have shown that the Wald interval also exhibits "erratic behavior" even when sample size is large.

An alternative approximate method that seems to have broad support in the statistical literature is the Wilson, or Score Test, interval (Brown et al 2001 & 2002, Agresti and Coull 1998). It is an inversion of the score test, and the development of the equation is reviewed in Agresti and Coull (1998). The Wilson interval is proposed in preference to the Wald interval for improving assessment of ambient-based standards. The equation for the Wilson interval is somewhat intimidating (Equation 5), but it is manageable on a spreadsheet.

$$\left\{\hat{p} + z_{\alpha/2}^2/2n \pm z_{\alpha/2}\sqrt{\left[\hat{p}(1-\hat{p}) + z_{\alpha/2}^2/4n\right]/n}\right\}/(1 + z_{\alpha/2}^2/n)$$

Equation 5. The Wilson confidence interval for an estimated percentile.

Adaptation of Wilson Interval Method

The Wilson interval method is an improvement over the exact method and the Wald interval, but the discreteness issue remains. In exact and approximate methods, the target proportion ($\hat{p}$ = k/n) is formally defined by the [integer] number of "successes" (k) relative to sample size (n). As mentioned previously, most sample sizes do not match exactly the proportion (e.g., 85%) used for assessment.

The idiosyncrasies of Colorado's assessment method prompt consideration of a departure from the usual statistical approach of scoring "successes" on the basis of individual samples. In current methodology, assessment is based on locating the concentration that corresponds to the 85th percentile of the assessed data set. Usually, this step is carried out with the PERCENTILE function in EXCEL, which interpolates between measured concentrations. In other words, the assessed value rarely matches a measured concentration from the assessed data set.

APPENDIX B

The equation for the Wilson interval requires values for $\hat{p}$, n, and z. There is no computational impediment to finding the LCL for any percentile in any data set. Hence, the equation is used to define the LCL for the three conventionally-assessed percentiles (50^{th}, 85^{th}, and 95^{th}) over a wide range of possible sample sizes (n = 5 through 100) with a one-sided, 95% confidence level (i.e., $\alpha = 0.05$). The values are summarized in **Supporting Table 1**. However, due to a quirk in the way EXCEL calculates percentiles, the values in **Supporting Table 1** must be adjusted, as explained later, when assessments are undertaken with EXCEL spreadsheets.

Translation of the LCL into an Excel-compatible Percentile

Percentiles in Excel are calculated by a method that does not match the calculations used to establish the LCLs. The computational differences are not major, but become increasingly important for small sample sizes. Insofar as the Excel function is widely used, we have incorporated an adjustment such that **Supporting Table 2** contains Excel-compatible LCLs. The basis for the adjustment and the interpretation of **Supporting Table 2** are given below.

Percentiles in Excel are set such that the smallest value is defined as 0% and the largest value is defined as 100%. The formula is p' = (k-1)/(N-1), where k is the rank of the observation and N is the number of observations in the data set (see Schoonjans et al 2011). Development of the LCL is based on the binomial distribution, which defines percentiles for each ranked observation as k/N. The largest concentration is still set to 100%, but the smallest observed concentration is 1/N rather than 0%. Thus, when the LCL is converted to a concentration, it could be smaller than the smallest observed value (and thus represent a percentile between zero and 1/N).

The difference between the two formulas affects conversion of the LCL to a concentration using EXCEL functions. Direct conversion of the LCL from **Supporting Table 1** with an Excel function would not yield the correct value. The error is very small when sample size is large, but cannot be ignored at small sample sizes. Therefore, an adjustment should be made.

The adjustment relies on simple algebra to translate the LCL from **Supporting Table 1** into LCL' that is compatible with Excel functions. In the Excel formula, where LCL' = (k-1)/(N-1), the binomial formula, LCL = k/N, is rearranged to enable substitution for k (k=LCL*N). After substituting for k, the equation for the translation is LCL' = (LCL*N-1)/(N-1). This translation is used to produce an Excel-compatible set of LCLs in **Supporting Table 2**. The threshold concentration can be derived directly with the Excel function as follows, PERCENTILE(*concentration_data*,LCL') where the assessed concentration data are in a defined range of cells.

Comparison to Exact Method

Exact and approximate methods can be compared only in those few cases where the integer values of k yield the target percentiles because the proposed method for locating the LCL is not restricted to discrete values of k (p=k/n). For the 50^{th} percentile, the comparison can be made for all even values of n; it shows that the exact method is more conservative, as expected, especially at small sample size. The exact and approximate methods agree within

10% when n exceeds about 20 (Figure 4.). Agreement between the two methods is generally even better for the 85th and 95th percentiles than the 50th percentile.

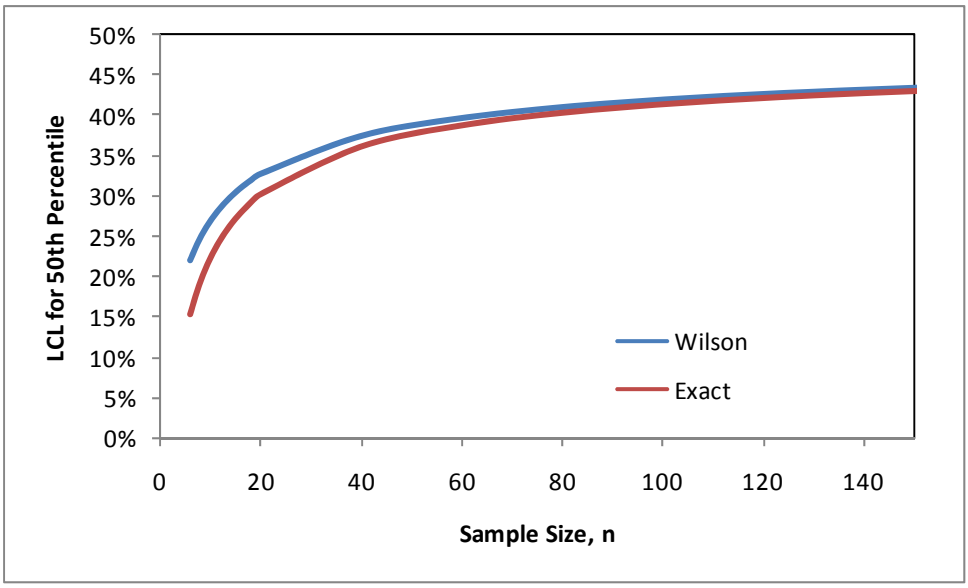


Figure 4. Comparison of LCLs defined by Wilson and exact methods for the 50th percentile. Comparison is limited to sample sizes with even numbers because the exact method can be evaluated only with integer values.

Sample Size

Sample sizes for water quality assessments tend to be small, often less than 10 samples for a segment. As is evident from **Supporting Table 1**, confidence intervals for percentiles are broader when sample sizes are small. Although LCL values can be calculated with the Wilson interval equation for virtually any sample size, there is a practical reason for avoiding very small sample sizes. When the percentile of the LCL is smaller than that of the lowest measured concentration (i.e., <1/N), it would correspond to a concentration smaller than any that were measured. For example, when only four samples are available, the LCL for the 50th percentile would be 0.182, which is smaller than 1/N (=0.25=1/4). When five samples are available, the LCL for the 50th percentile is 0.204, and this is larger than 1/N (=0.200). For the purpose of making listing decisions with ambient-based standards, at least five samples are required.

Setting a minimum of five samples for assessment of ambient-based standards differs from current assessment practice. With the latest version of Colorado's assessment methodology, a firm decision to place a segment on the 303(d) list requires more than ten samples (assuming the decision is based solely on concentration data for one constituent). With 4 to 10 samples, a listing decision must be backed up with evidence in addition to measured concentrations. For three or fewer samples, high concentrations would at most trigger further sampling (M&E list).

Adoption of the LCL table could simplify decision-making for the listing methodology by having only two pathways related to sample size. In order to make a listing decision, there

must be at least five representative[8] samples. When there are at least five samples, no additional supporting information is required because conclusions are equally reliable whether sample size is five or ten or fifty. When there are less than five representative samples, no action should be taken. If there is insistence on having a pathway to the M&E list, it should be on the basis of a table with a smaller confidence interval (e.g., 90%), which would also include a lower threshold for sample size.

Examples with Assessment Data

The current methodology for assessing attainment of ambient-based standards can be improved substantially by adding a defined level of confidence for the attainment decision. The statistical justification for the change is strong. Working through examples with real data is a good way to show that the improved approach is practical and efficient. The following examples incorporate data from the historical record for illustrative purposes, and they are not intended for reaching conclusions in the formal assessment process. Standards or segment descriptions may be changed through Commission actions (as happened with segment COARMA04a), and more recent data may be available for assessment. Nevertheless, these examples retain value for comparing old and new assessment methods for ambient-based standards.

One example is taken from Wildhorse Creek above Highway 50, which has ambient-based standards for selenium. Recent measurements (N=36; 2005-2011) of selenium concentrations serve as the assessed data set (Table 2). Most of the observed concentrations exceed table value standards (ch=4.6 ug/L; ac=18.4 ug/L) by a wide margin.

Date	Se, ug/L	Date	Se, ug/L	Date	Se, ug/L
10/13/2005	420	4/3/2007	500	7/1/2010	441
12/19/2005	496	6/5/2007	429	8/1/2010	564
2/15/2006	593	6/13/2007	95	9/1/2010	487
3/2/2006	535	9/20/2007	754	10/1/2010	479
4/26/2006	480	12/18/2007	691	11/1/2010	539
5/31/2006	362	12/1/2009	556	11/16/2010	1900
7/6/2006	9	1/1/2010	355	12/1/2010	618
9/5/2006	410	2/1/2010	646	2/1/2011	554
10/11/2006	361	3/1/2010	641	2/7/2011	1800
12/5/2006	224	4/1/2010	728	3/1/2011	607
2/20/2007	340	5/1/2010	605	4/1/2011	581
3/9/2007	531	6/1/2010	536	5/24/2011	1500

Table 2. Selenium concentrations measured in Wildhorse Creek above Highway 50.

The first step in assessing ambient-based standards by current methodology is to determine the 85th and 95th percentiles[9] of concentrations in the assessed data set. The 85th percentile

[8] The most recent listing methodology describes the factors to be considered when judging if data are "representative". Factors typically include spatial and temporal distribution of sampling effort, as well as analytical considerations and atypical events in the watershed.

[9] Consistent with current methodology, threshold concentrations are determined with EXCEL's PERCENTILE function.

concentration of the assessed data set is 680 ug/L, which exceeds the ambient-based chronic standard of 597 ug/L (Table 3). The 95th percentile concentration is 1575 ug/L, which exceeds the ambient-based acute standard of 708 ug/L. The assessed values exceed the ambient-based standards, which, by current assessment methodology, would trigger a listing.

Metric	Chronic	Acute
Ambient-based Standard	597	708
Assessed 85th Percentile	680	---
Assessed 95th Percentile	---	1575
Current outcome	Exceeded	Exceeded
Sample Size	36	36
LCL percentile (Appendix A)	0.728	0.853
LCL, Excel-compatible (App B)	0.720	0.849
LCL concentration	605	678
New outcome	Exceeded	OK

Table 3. Assessment of data for attainment of ambient-based selenium standards in Wildhorse Creek. LCL percentiles are taken from Appendices with N=36. The ambient-based standards were appropriate for segment COARMA04a at the time the samples were taken, but changes were adopted subsequently by the Commission.

The selenium data from Wildhorse Creek also are assessed with the improved methodology. Based on the sample size of 36, percentiles for chronic and acute LCLs are taken from the columns corresponding to the 85th and 95th percentiles in **Supporting Table 2**. The LCL for the 85th percentile is 0.720 (Excel-compatible value), which corresponds to a selenium concentration of 605 ug/L in the assessed data set. Therefore, the 85th percentile of the assessed data set is significantly larger than the chronic standard, which would trigger a listing by the improved methodology.

The conclusion about attainment can be properly reached based on assessment of the chronic standard alone, but the data also are assessed for attainment of the acute standard for illustrative purposes. In this case, the LCL corresponds to a selenium concentration of 678 ug/L, which is less than the acute standard. If assessment had been based solely on the acute standard, the conclusion from the improved methodology would have been that the assessed value was not significantly larger than the acute standard, which would not trigger a listing.

A second example is taken from Big Dry Creek, which has a seasonal, ambient-based chronic standard for selenium; the acute standard remains equal to the table value standard (TVS). The example deals only with data from the Apr-Oct "season". Recent measurements (N=34; 2006-2010) of selenium concentrations serve as the assessed data set (Table 4). Some of the observed concentrations exceed the chronic TVS (4.6 ug/L), but none exceeds the acute TVS (18.4 ug/L).

Date	Se, ug/L	Date	Se, ug/L	Date	Se, ug/L
4/13/2006	8.2	6/19/2008	5.7	8/12/2010	2.3
4/12/2007	9.1	6/18/2009	2.9	9/14/2006	2.8
4/17/2008	6.4	6/10/2010	3.0	9/13/2007	2.9
4/9/2009	7.0	7/13/2006	5.5	9/11/2008	7.6
4/8/2010	5.7	7/12/2007	4.9	9/10/2009	3.7
5/11/2006	2.4	7/17/2008	2.3	9/9/2010	3.5
5/10/2007	3.4	7/9/2009	2.6	10/19/2006	5.6
5/8/2008	4.0	7/8/2010	2.3	10/11/2007	10.3
5/14/2009	7.5	8/10/2006	4.4	10/9/2008	8.1
5/13/2010	3.3	8/9/2007	7.1	10/15/2009	9.5
6/15/2006	2.9	8/14/2008	8.1		
6/14/2007	2.5	8/13/2009	5.3		

Table 4. Selenium concentrations measured in Big Dry Creek above the USGS gage (COSPBD01). Data are shown for Apr through Oct, 2006-2010.

When the selenium standards for Big Dry Creek are assessed by current methodology, the chronic standard is exceeded, but the acute standard is not (Table 5). The assessment would likely trigger a listing by current methodology. The data from Big Dry Creek also are assessed for the chronic standard with the improved methodology. The LCL for the 85th percentile is 0.716, which corresponds to a selenium concentration of 6.8 ug/L in the assessed data set. Therefore, the 85th percentile of the assessed data set is not significantly larger than the standard and would not trigger a listing. Application of the improved methodology highlights the value of using a defined level of confidence to support listing decisions.

Metric	Chronic	Acute
Ambient-based Standard	7.4	18.4
Assessed 85th Percentile	8.1	---
Assessed 95th Percentile	---	9.3
Current outcome	Exceeded	OK
Sample Size	34	
LCL percentile, Excel compatible	0.716	
LCL concentration	6.8	
New outcome	OK	

Table 5. Assessment of data for attainment of selenium standards in Big Dry Creek. The chronic standard is ambient-based, and the acute standard is TVS. Consequently, only the chronic standard is assessed with the improved methodology. The LCL percentile is taken from Appendix B with N=34.

General Comments

The proposed addition to assessment methodology for ambient-based standards offers an important statistical improvement by establishing a defined level of confidence to support impairment decisions. Obtaining the benefit of a defined confidence level comes at little additional cost because assessment requires almost no additional effort. In addition, the improved methodology retains the practical advantages of the current methodology in that it works for any constituent and requires no assumptions about the underlying statistical distribution.

APPENDIX B

The approach developed by the Division is efficient and effective for routine application, and the Division plans to apply it in the next listing cycle. Nevertheless, it is not the only approach that can provide a defined level of confidence. For example, a bootstrap approach could provide a non-parametric basis for assessment. Where a distribution can be identified, parametric options could be employed. These alternatives may be suitable where the resources are available to invest in developing a statistically-defensible approach for a particular water body.

Improving confidence in assessment decisions by any statistically-defensible method addresses important concerns about future commitment of resources where impairment is identified. By incorporating a defined level of confidence in assessments of ambient-based standards, the Division can be more certain that resources committed to TMDL development, for example, will not be wasted. A more reliable basis for listing decisions also should be well-received by stakeholders, who are affected by listing decisions (or reversals).

Increasing the statistical rigor of assessment also creates more incentive to set minimum requirements for development of ambient-based standards. Sample size and representativeness merit discussion that is beyond the scope of this assessment methodology.

Development of an improved approach for ambient-based standards invites the question of why a similar approach is not also proposed for TVS assessments. In the Division's view, there are important differences between the two kinds of standards. A TVS generally represents a physiological threshold above which concentrations threaten aquatic life. In contrast, assessment of ambient-based standards hinges on detecting degradation of water quality: Have concentrations increased significantly over "existing quality"? Thus, the Division does not recommend a change in current assessment practice for TVS.

References

Agresti, A and BA Coull. 1998. Approximate is better than "exact" for interval estimation of binomial proportions. The American Statistician 52(2): 119-126.

Brown, LD, TT Cai, and A DasGupta. 2001. Interval estimation for a binomial proportion. Statistical Science 16(2): 101-133 (includes appended commentary by several authors).

Brown, LD, TT Cai, and A DasGupta. 2002. Confidence intervals for a binomial proportion and asymptotic expansions. The Annals of Statistics 30(1): 160-201.

Helsel, DR, and RM Hirsch. 2002. Statistical methods in water resources. Chapter A3, Book 4 Hydrologic Analysis and Interpretation. Techniques of Water-Resources Investigations of the United States Geological Survey. http://water.usgs.gov/pubs/twri/twri4a3/

Leemis, LM and KS Trivedi. 1996. A comparison of approximate interval estimators for the Bernoulli parameter. The American Statistician 50: 63-68.

McBride, GB. 2005. Using statistical methods for water quality methods. Wiley-Interscience. 313 p.

APPENDIX B

Schoonjans, F, D De Bacquer, and P Schmid. 2011. Estimation of population percentiles. Epidemiology 22(5): 750-751.

Zar, JH. 1999. Biostatistical Analysis, 4th Edition. Prentice-Hall. 663 p. + Appendices.

APPENDIX B

Supporting Table 1: Lower confidence limits (LCLs) for three percentiles of regulatory interest (50th, 85th, and 95th) at assessed sample sizes of 5 through 100. These percentiles should not be used with the PERCENTILE function in EXCEL; the EXCEL-compatible values are presented in Appendix B.

N	$\hat{p}$ = 0.50	$\hat{p}$ = 0.85	$\hat{p}$ = 0.95		N	$\hat{p}$ = 0.50	$\hat{p}$ = 0.85	$\hat{p}$ = 0.95
5	0.204	0.482	0.588		53	0.390	0.752	0.875
6	0.221	0.514	0.625		54	0.391	0.754	0.876
7	0.236	0.540	0.654		55	0.392	0.755	0.877
8	0.249	0.561	0.678		56	0.393	0.756	0.878
9	0.260	0.579	0.698		57	0.394	0.756	0.879
10	0.269	0.595	0.715		58	0.394	0.757	0.880
11	0.278	0.608	0.730		59	0.395	0.758	0.881
12	0.286	0.619	0.742		60	0.396	0.759	0.881
13	0.292	0.630	0.753		61	0.397	0.760	0.882
14	0.299	0.639	0.763		62	0.398	0.761	0.883
15	0.305	0.647	0.772		63	0.399	0.762	0.884
16	0.310	0.654	0.780		64	0.399	0.763	0.884
17	0.315	0.661	0.787		65	0.400	0.763	0.885
18	0.319	0.667	0.793		66	0.401	0.764	0.886
19	0.323	0.673	0.799		67	0.401	0.765	0.886
20	0.327	0.678	0.804		68	0.402	0.765	0.887
21	0.331	0.683	0.809		69	0.403	0.766	0.887
22	0.335	0.687	0.813		70	0.404	0.767	0.888
23	0.338	0.692	0.818		71	0.404	0.768	0.889
24	0.341	0.695	0.821		72	0.405	0.768	0.889
25	0.344	0.699	0.825		73	0.405	0.769	0.890
26	0.346	0.703	0.828		74	0.406	0.769	0.890
27	0.349	0.706	0.832		75	0.407	0.770	0.891
28	0.352	0.709	0.834		76	0.407	0.771	0.891
29	0.354	0.712	0.837		77	0.408	0.771	0.892
30	0.356	0.714	0.840		78	0.408	0.772	0.892
31	0.358	0.717	0.842		79	0.409	0.772	0.893
32	0.360	0.719	0.845		80	0.410	0.773	0.893
33	0.362	0.722	0.847		81	0.410	0.774	0.894
34	0.364	0.724	0.849		82	0.411	0.774	0.894
35	0.366	0.726	0.851		83	0.411	0.775	0.895
36	0.368	0.728	0.853		84	0.412	0.775	0.895
37	0.369	0.730	0.855		85	0.412	0.776	0.895
38	0.371	0.732	0.856		86	0.413	0.776	0.896
39	0.373	0.734	0.858		87	0.413	0.777	0.896
40	0.374	0.735	0.860		88	0.414	0.777	0.897
41	0.376	0.737	0.861		89	0.414	0.777	0.897
42	0.377	0.738	0.863		90	0.415	0.778	0.897
43	0.378	0.740	0.864		91	0.415	0.778	0.898
44	0.380	0.741	0.865		92	0.415	0.779	0.898
45	0.381	0.743	0.867		93	0.416	0.779	0.898
46	0.382	0.744	0.868		94	0.416	0.780	0.899
47	0.383	0.745	0.869		95	0.417	0.780	0.899
48	0.385	0.747	0.870		96	0.417	0.781	0.900
49	0.386	0.748	0.871		97	0.418	0.781	0.900
50	0.387	0.749	0.872		98	0.418	0.781	0.900
51	0.388	0.750	0.873		99	0.418	0.782	0.901
52	0.389	0.751	0.874		100	0.419	0.782	0.901

APPENDIX B

Supporting Table 2: Lower confidence limits (LCLs) for three percentiles of regulatory interest (50th, 85th, and 95th) at assessed sample sizes of 5 through 100. These percentiles are compatible with the PERCENTILE function in EXCEL; see text for explanation.

N	$\hat{p}$ = 0.50	$\hat{p}$ = 0.85	$\hat{p}$ = 0.95		N	$\hat{p}$ = 0.50	$\hat{p}$ = 0.85	$\hat{p}$ = 0.95
5	0.005	0.353	0.485		53	0.378	0.748	0.873
6	0.066	0.417	0.550		54	0.379	0.749	0.874
7	0.109	0.464	0.597		55	0.380	0.750	0.875
8	0.141	0.499	0.632		56	0.382	0.751	0.876
9	0.167	0.527	0.661		57	0.383	0.752	0.877
10	0.188	0.550	0.684		58	0.384	0.753	0.878
11	0.206	0.569	0.703		59	0.385	0.754	0.878
12	0.221	0.585	0.719		60	0.386	0.755	0.879
13	0.234	0.599	0.733		61	0.387	0.756	0.880
14	0.245	0.611	0.745		62	0.388	0.757	0.881
15	0.255	0.622	0.755		63	0.389	0.758	0.882
16	0.264	0.631	0.765		64	0.390	0.759	0.882
17	0.272	0.640	0.773		65	0.391	0.760	0.883
18	0.279	0.648	0.781		66	0.392	0.760	0.884
19	0.286	0.655	0.788		67	0.392	0.761	0.884
20	0.292	0.661	0.794		68	0.393	0.762	0.885
21	0.298	0.667	0.799		69	0.394	0.763	0.886
22	0.303	0.673	0.804		70	0.395	0.763	0.886
23	0.308	0.678	0.809		71	0.396	0.764	0.887
24	0.312	0.682	0.814		72	0.396	0.765	0.888
25	0.316	0.687	0.818		73	0.397	0.766	0.888
26	0.320	0.691	0.822		74	0.398	0.766	0.889
27	0.324	0.694	0.825		75	0.399	0.767	0.889
28	0.328	0.698	0.828		76	0.399	0.768	0.890
29	0.331	0.701	0.831		77	0.400	0.768	0.890
30	0.334	0.704	0.834		78	0.401	0.769	0.891
31	0.337	0.707	0.837		79	0.401	0.769	0.891
32	0.340	0.710	0.840		80	0.402	0.770	0.892
33	0.342	0.713	0.842		81	0.403	0.771	0.892
34	0.345	0.716	0.844		82	0.403	0.771	0.893
35	0.347	0.718	0.847		83	0.404	0.772	0.893
36	0.350	0.720	0.849		84	0.405	0.772	0.894
37	0.352	0.722	0.851		85	0.405	0.773	0.894
38	0.354	0.725	0.853		86	0.406	0.773	0.895
39	0.356	0.727	0.854		87	0.406	0.774	0.895
40	0.358	0.728	0.856		88	0.407	0.774	0.895
41	0.360	0.730	0.858		89	0.407	0.775	0.896
42	0.362	0.732	0.859		90	0.408	0.775	0.896
43	0.364	0.734	0.861		91	0.409	0.776	0.897
44	0.365	0.735	0.862		92	0.409	0.776	0.897
45	0.367	0.737	0.864		93	0.410	0.777	0.897
46	0.368	0.738	0.865		94	0.410	0.777	0.898
47	0.370	0.740	0.866		95	0.411	0.778	0.898
48	0.371	0.741	0.867		96	0.411	0.778	0.898
49	0.373	0.743	0.869		97	0.412	0.779	0.899
50	0.374	0.744	0.870		98	0.412	0.779	0.899
51	0.376	0.745	0.871		99	0.413	0.779	0.900
52	0.377	0.747	0.872		100	0.413	0.780	0.900