4. Using Chemical Data as Indicators of Water Quality

Contents

4.1 How Are Chemical Data Used Within the Context of the State's WQS? 4-2

4.1.1 Numeric Criteria 4-3

4.1.2 Narrative Criteria 4-7

4.2 What Actions Does the State Take to Assess and Document Data Quality, Including Third-Party Data? 4-8

4.2.1 How Does the State Define Data Quality? 4-9

4.2.2 How Does the State Assess (Review and Evaluate) Data Quality? 4-12

4.2.3 How Does the State Document the Level of Data Quality? 4-13

4.3 How Does the State Analyze and Interpret Chemical Data To Determine WQS Attainment/Impairment? 4-13

4.3.1 What Statistical Analyses for Interpreting Chemical Data Does the State Use? 4-18

4.3.2 How Does the State Make Attainment/Impairment Decisions in the Absence of a "Perfect Data Set"? 4-20

4.4 References 4-20

4. Using Chemical Data as Indicators of Water Quality

Complete assessment of water quality demands consideration of different types of data because each type provides unique insights into water quality standards (WQS) attainment status. This chapter addresses the role of chemical data in assessing WQS attainment status and listing impaired waters. Subsequent chapters cover biological, toxicity, bacteria, and habitat data, respectively. Note that conventional indicators such as temperature, pH, and dissolved oxygen, which are sometimes referred to as physical data, are included in this chapter because they are generally treated as chemical indicators of water quality.

Chemical data are important indicators of WQS attainment/impairment for a number of reasons. All state, territory, and authorized tribal WQS include chemical-specific numeric water quality criteria adopted to protect aquatic life and human health from the effects of pollution. Assessments of chemical concentrations serve as direct measures of stressors to aquatic life and human health. Chemical-specific data and water quality models allow predictions of the likelihood of impacts to aquatic life and human health where they may not yet have occurred. Chemical pollutants also lend themselves to chemical-specific total maximum daily limit (TMDL) development and source controls, particularly as expressed in National Pollution Discharge Elimination System (NPDES) discharge permits.

Using chemical data involves issues related to data quality as well as ensuring that data are representative of water quality conditions. This chapter helps states reduce uncertainty by documenting their approaches for using chemical data to make WQS attainment decisions and list impaired waters. Each section title poses a question that addresses an element of a state's assessment and listing methodology.

4.1 How Are Chemical Data Used Within the Context of the State's WQS?

State WQS play a central role in a state's water quality management program. Standards drive water quality assessments, 303(d) lists of impaired waters, 305(b) reports on water quality status and trends, TMDLs, NPDES permits, and nonpoint-source management measures. These standards include designated uses appropriate for each waterbody, numeric and narrative criteria adopted to protect uses, and policies to prevent degradation of waters. Chemical data primarily support assessments of the extent to which numeric and narrative criteria are met. The state's assessment and listing methodology should describe how chemical data are collected and how they are used to determine attainment of WQS.

States, territories, and authorized tribes adopt water quality criteria to protect designated uses, including aquatic life, recreation, public water supplies, and fish and shellfish consumption. The criteria should be based on sound scientific rationale and should contain sufficient indicators or parameters to protect the designated uses. Water quality criteria are numeric criteria derived from EPA's 304(a) criteria guidance documents or other scientifically defensible methods, or narrative criteria adopted when numeric criteria cannot be determined or to supplement numeric criteria.

4.1.1 Numeric Criteria

Under section 304 of the CWA, states, territories, and authorized tribes adopt chemical-specific numeric criteria into their WQS to protect designated uses. These criteria generally include:

- Aquatic life thresholds for acute or chronic exposure of sensitive organisms,

- Human health thresholds for cancer risk or noncancer risk due to exposure via drinking water and fish tissue consumption, and

- Organoleptic effect thresholds for drinking water consumption and recreation.

The Final Water Quality Guidance for the Great Lakes System outlines a process for developing numeric criteria to protect wildlife (U.S. EPA 1995).

EPA, under section 304(a) of the CWA, periodically publishes recommendations (guidance) for use by states and authorized tribes in developing and adopting criteria protective of designated uses. In adopting such criteria, states and tribes may use (1) 304(a) criteria, (2) 304(a) criteria modified to reflect site specific conditions, or (3) other scientifically defensible methods (see 40 CFR 131.11). The complete listing of EPA-recommended water quality criteria can be found in National Recommended Water Quality Criteria–Correction (EPA 822-Z-99-001) or the most recent update thereof at http://www.epa.gov/ost/standards/wqcriteria.html. Table 4-1 lists state websites where individual state WQS, including numeric criteria, are presented in detail. A source for the effective state water quality standards and criteria is on the EPA website at: http://www.epa.gov/wqsdatabase.

Numeric criteria for aquatic life protection

Development of numeric water quality criteria for aquatic life protection is a complex process described in the 1985 Guidelines and in EPA's criteria guidance documents and summarized in the Water Quality Standards Handbook (U.S. EPA 1994). The process involves collecting and analyzing data on a specific chemical concerning its toxicity to aquatic organisms. To serve as a basis for criteria development, data must be available for at least one species in each of at least eight different families. If sufficient data are available, EPA derives a recommended acute (criteria maximum concentration, CMC) and chronic (criteria continuous concentration, CCC) criterion. Acute thresholds estimate the highest 1-hour concentration that will not have an unacceptable lethal effect on 95% of the species tested. Similarly, chronic thresholds estimate the highest 4-dayconcentration that should not cause unacceptable toxicity during long-term exposure. Acute or chronic criteria can be adjusted to reflect water quality characteristics such as pH, temperature, or hardness. Separate criteria may be developed for fresh and salt waters.

Table 4-1. State agency web sites for water quality standards and criteria (current as of February 2001)

State	WQS Web Address
AK	http://www.state.ak.us/local/akpages/ENV.CONSERV/dawq/dm/wqsmain/regs.htm
AL	http://www.adem.state.al.us/RegsPermit/ADEMRegs/Div6Vol1/rdiv6v1.html http://www.adem.state.al.us/RegsPermit/PropRules/proprule.htm
AR	http://www.adeq.state.ar.us/regs/reg02.htm
AZ	http://www.sosaz.com/public_services/Title_18/18-11.htm
CA	http://www.swrcb.ca.gov/plnspols/index.html
CO	http://www.cdphe.state.co.us/cdphereg.asp#wqreg
CT	http://dep.state.ct.us/wtr/wqs.pdf
DE	http://www.dnrec.state.De.us/water/wqs1999.pdf
FL	http://www.dep.state.fl.us/ogc/documents/rules/shared/62-302.pdf http://www.dep.state.fl.us/ogc/documents/rules/shared/62-302t.pdf
GA	http://www.ganet.org/dnr/environ/rules_files/exist_files/391-3-6.pdf
HI	http://mano.icsd.hawaii.gov/doh/rules/ADMRULES.html
IA	http://web.legis.state.ia.us/Rules/2000/iac/567iac/56761/
ID	http://www2.state.id.us/adm/adminrules/rules/IDAPA58/58INDEX.HTM http://www2.state.id.us/adm/adminrules/bulletin/sept00.pdf
IL	http://www.ipcb.state.il.us/title35/download/C302.pdf
IN	http://www.ai.org/legislative/iac/title327.html
KS	http://www.kdhe.state.ks.us/download/index.html#bowreports
KY	http://www.lrc.state.ky.us/kar/401/005/026.htm
LA	http://www.deq.state.la.us/planning/regs/title33/index.htm#partix
MA	http://www.state.ma.us/dep/brp/wm/files/314cmr4.pdf
MD	http://209.15.49.5/dsd_web/default.htm
ME	http://janus.state.me.us/legis/statutes/38/title38ch30sec0.html
MI	http://www.deq.state.mi.us/swq/
MN	http://www.revisor.leg.state.mn.us/arule/7050/
MO	http://mosl.sos.state.mo.us/csr/10csr/10c20-7.pdf
MS	http://www.deq.state.ms.us/newweb/opchome.nsf/pages/SurfaceWaterfiles/$file/wqc.pdf
MT	http://www.deq.state.mt.us/dir/Legal/Chapters/CH30-06.pdf
NC	http://mapsweb01.sips.state.nc.us/ncoah/ncadministrativ_/title15aenviron_/chapter02enviro_/default.htm
ND	N/A
NE	http://www.deq.state.ne.us/RuleAndR.nsf/pages/117-TOC
NH	http://www.des.state.nh.us/wmb/Env-Ws1700.pdf
NJ	http://www.state.nj.us/dep/landuse/njac/7-9b.pdf http://www.state.nj.us/dep/watershedmgt/swqs/

Table 4-1. State agency web sites for water quality standards and criteria (current as of February 2001) (continued)

State	WQS Web Address
NM	http://www.nmenv.state.nm.us/NMED_regs/swqb/20nmac6_1.html
NV	http://www.leg.state.nv.us/NAC/NAC-445A.html
NY	http://www.dec.state.ny.us/website/regs/ch10.htm
OH	http://www.epa.state.oh.us/dsw/rules/3745-1.html
OK	http://www.state.ok.us/~orwb/rules/Chap45.pdf
OR	http://waterquality.deq.state.or.us/wq/wqrules/wqrules.htm
PA	http://www.pacode.com/secure/data/025/chapter93/chap93toc.html
RI	http://www.state.ri.us/dem/REGS/WATER/QUALREGS.PDF
SC	http://www.scdhec.net/water
SD	http://legis.state.sd.us/rules/rules/7451.htm
TN	http://www.state.tn.us/sos/rules/1200/1200-04/1200-04.htm
TX	http://www.tnrcc.state.tx.us/oprd/rules/pdflib/307`.pdf
UT	http://www.rules.state.ut.us/publicat/code/r317/r317-002.htm
VA	http://ftp.deq.state.va.us/pub/watrregs/wqs.zip
VT	http://www.state.vt.us/wtrboard/july2000wqs.htm
WA	http://www.ecy.wa.gov/biblio/wac173201a.html
WI	http://www.legis.state.wi.us/rsb/code/nr/nr100.html
WV	http://www.state.wv.us/csr/docs/WPDocs/4601 .wpd

Note: N/A means WQS not on the web or web address not available at time of compilation.

EPA's criteria guidelines for aquatic life protection recommend that a criterion is comprised of a chemical concentration, a duration, and a frequency. The acute criterion (criteria maximum concentration, CMC) equals the highest concentration of a pollutant to which the aquatic species can be exposed for a short period of time without deleterious effects. The chronic criterion (criteria continuous concentration, CCC) equals the highest concentration of a pollutant to which the aquatic species can be exposed for an extended period of time (4 days) without deleterious effects. For ammonia, a 30-day rather than a 4-day average is recommended. Alternative averaging periods can be developed by using data that relate toxic response with exposure time, or by using models of toxicant uptake and action (U.S. EPA 1991b). Both the acute and chronic exposure durations were set to be fully protective of fast-acting toxicants, and are therefore even more protective for slower acting toxicants.

Early in the WQS program, EPA criteria guidance for several parameters, including chlorides, turbidity, and temperature, stated that these criteria should not be exceeded at any frequency. Later EPA guidance distinguished between conventional pollutants and toxic pollutants when providing recommendations about the number of exceedances that constitute nonattainment of WQS. For conventional pollutants, the 305(b) guidelines indicated that whenever more than 10% of the water quality samples collected exceed the criterion threshold, the WQS is not attained (U.S. EPA 1997).

EPA recommended that acute and chronic aquatic life criteria for toxics not be exceeded more than once every 3-year period on the average. EPA selected this frequency to provide a level of protection similar to the 7Q10 design flow or low-flow condition. The exceedance frequency recommendation is considered protective. Like the magnitude and duration components of the water quality criteria, it may also be revised to reflect site-specific information on exposure and response relationships.

Numeric criteria for human health protection

States adopt numeric chemical criteria for human health protection as part of WQS developed to protect public water supply, fish consumption, and recreational uses of surface waters. States may adopt numeric fish-tissue-based chemical criteria for the protection of human health from consumption of contaminants such as mercury in fish.

In 2000, EPA published revisions to the methodology for developing water quality criteria for the protection of human health (U.S. EPA 2000c). These revisions incorporate the latest scientific information for developing water quality criteria, including systematic procedures for evaluating cancer risk, noncancer health effects, human exposure, and bioaccumulation potential in fish. (See also http://www.epa.gov/ost/humanhealth/method/index.html.)

The revised methodology provides more flexibility for decision making at state, tribal, and EPA regional levels. Specifically, it provides opportunities for states, territories, and tribes to use tailored information on fish consumption rates, acceptable risk levels, and other factors that influence the calculations of chemical criteria. EPA believes that adoption of water quality criteria requires several risk management decisions that are often better made at the state, territory, or tribal level.

Water quality criteria to protect human health are generally based on protecting against long-term exposure to low concentrations of a toxic pollutant. When a chemical human health criterion is applied to WQS attainment decisions, EPA recommends evaluating comparing the mean (or geometric mean if appropriate for a skewed data set) of the measured concentrations with the criterion. However, some states have adopted human-health-based chemical criteria that establish instantaneous maximum concentrations, for which any exceedance constitutes nonattainment. If the mean or geometric mean exceeds the criterion, the WQS is not being attained.

4.1.2 Narrative Criteria

To supplement numeric criteria for toxic chemicals, states adopt narrative criteria. These criteria help ensure that all designated uses are protected under a wide range of circumstances. Narrative criteria are effective tools for addressing toxic effects of pollutants, exposure pathways, or exposure conditions for which the state has not adopted chemical-specific numeric criteria. Recommended narrative criteria, which are often referred to as "free froms," were first developed in 1968 and continue to be an important element of state, territorial, and tribal WQS.

EPA guidance explains that these "free froms" should apply to all waters of the United States at all flow conditions, including ephemeral and intermittent streams (U.S. EPA 1994). Narrative criteria guidance indicates that all waters should be free from substances that:

- Cause injury to, are toxic to, or produce adverse physiological responses in humans, animals, or plants
- Settle to form objectionable deposits
- Float as debris, scum, oil, or other material in concentrations that form nuisances
- Produce objectionable color, odor, taste, or turbidity
- Produce undesirable aquatic life or result in the dominance of nuisance species

States, territories, and authorized tribes may use chemical data are used to interpret a narrative criterion. For example, a state may use chemical concentrations in sediment, in conjunction with other information on sediment toxicity and the health of benthic communities, to identify a water as impaired because of sediment contamination. Another example is the use of fish tissue data. The concentrations of pollutants in fish tissue can be used in risk-based calculations to assess attainment of the fish consumption use as well as to issue fish consumption advisories. States may use narrative criteria to determine that a surface water is impaired for its public water supply use. This decision might be triggered by a finding that a drinking water utility has violated a chemical-specific maximum contaminant level for treated water and that the chemical is present in the surface water.

EPA recently issued guidance on the use of fish and shellfish consumption advisories and certain shellfish growing area classifications in determining attainment of water quality standards and listing impaired waterbodies under section 303(d) of the Clean Water Act. This guidance also recommends (in part) that in instances where tissue concentrations of pollutants do not indicate an exceedance, states, territories, and authorized tribes translate the applicable narrative criteria on a site-specific basis or adopt site specific numeric criteria to account for higher than expected exposures from contaminated fish or shellfish tissue and protect designated uses. This is discussed in an October 24, 2000, letter issued by EPA's Office of Science and Technology and Office of Wetlands, Oceans and Watersheds (U.S. EPA 2000d).

EPA encourages states, territories, and authorized tribes to use chemical data to interpret narrative criteria; however, these jurisdictions should develop implementation procedures, often referred to as translators, that explain how different types of chemical data are used to make attainment/impairment decisions based on narrative criteria. These implementation procedures should be made available for review and comment by the public.

4.2 What Actions Does the State Take To Assess and Document Data Quality, Including Third-Party Data?

This is an important question because it acknowledges that all data may not be of equal value for assessing WQS attainment/impairment. Results of analyses of chemical data or any other type of data may be of limited value unless they are accompanied by documentation about sample collection, analytical methods, and quality control (QC) protocols. Poorly documented monitoring results may indicate potential problems, corroborate other data and information, or trigger additional monitoring, but they may not to provide an adequate basis of an attainment or impairment decision if they fail to meet accepted data quality requirements. Chemical data with good data-quality documentation should be used to support an attainment/impairment decision.

Several states are reexamining and better defining requirements for acceptable data and protocols for screening data adequacy prior to interpreting data to make WQS attainment decisions. EPA has extensive technical documents on this topic, some of which are listed in the references to this chapter. Documenting data quality requirements and data evaluation procedures is a critical element that states need to address in their WQS implementation procedures.

It is important to balance data quality requirements with common sense. Data quality requirements must be objective and inclusive. States, territories, and authorized tribes must consider all existing and readily available data when making WQS attainment/impairment decisions. For example, if a state shares a waterbody with another state, it must consider existing and readily available data from the state that shares the waterbody. Data should not be excluded solely because of their source or their age, without a reasonable explanation as to why they do not represent water quality conditions. Similarly, data collected using methods different from those the state prefers should be considered if the detection limits for the method are appropriate for both the criteria threshold and the concentration detected.

Whatever data collection methods are employed, the state should ensure that their sampling design program be implemented in such a way that each sample represents the variable conditions extant in the target water(s) at the locality, during the time period for which the WQS attainment/impairment assessment is intended. Probability-based sampling designs are one way of ensuring representativeness of the sample, and permit estimation of the uncertainty associated with the sample-based estimates of the means, proportions, and other statistics required for comparison to criterion values. When these conditions are met, procedures for formal statistical inference can be applied to make scientifically defensible WQS decisions (see Appendices C and D; Peterson et al. 1999).

4.2.1 How Does the State Define Data Quality?

EPA encourages states, territories, interstate commissions, and authorized tribes to use the data quality objectives process to define minimum data quality requirements. This includes information on appropriate sample size and monitoring design, sample collection and handling protocols, analytical methods and detection limits, QC procedures, and data management. Frequently this type of information is documented in the state's quality assurance (QA) project plan or standard operating procedures (SOPs) for monitoring. Data quality requirements may be defined in the applicable WQS or in other implementating regulations or policy and procedures documents. For example, the WQS may define critical conditions, such as flow or temperature, under which the criteria apply or should be modified, while the implementation procedures may discuss information like data quality objectives, samples sizes, and SOPs.

It is important to make this information available to other organizations such as tribal, interstate, state, Federal, academic, and volunteer citizen groups that also monitor water quality. Over time, these potential partners may agree to meet data quality requirements if an agency clearly spells out these requirements in its assessment and listing methodology or other readily available and well-publicized documents.

Sample size is an important element of data quality. In general, statistical tests have good power for detecting exceedances if they are based on data from samples composed of 30 or more sampling units. Smaller sample sizes are prone to yield erroneous attainment decisions because they have a low probability of detecting WQS exceedances unless they are large and pervasive. Figure 4-1 illustrates the effect of sample size on the probability of detecting exceedances when the actual proportion of exceedant sampling units exceeds the criterion proportion (e.g., 10% exceedance) by amounts between zero and 50%. Computing such power curves for different sample sizes can be an effective tool for illustrating the benefits of larger sample sizes. Similarly, computing the width of the confidence intervals for the exceedance proportion provides an indication of the uncertainty of the sample estimate for distinguishing impairment from attainment (Figure 4-2). Appendices C and D describes these issues in more detail and provides guidance and additional references on determining sample sizes.

If a state is uncomfortable basing attainment/impairment decisions on small data sets, it should demonstrate a commitment to collection of sufficient data to support its data quality objectives.

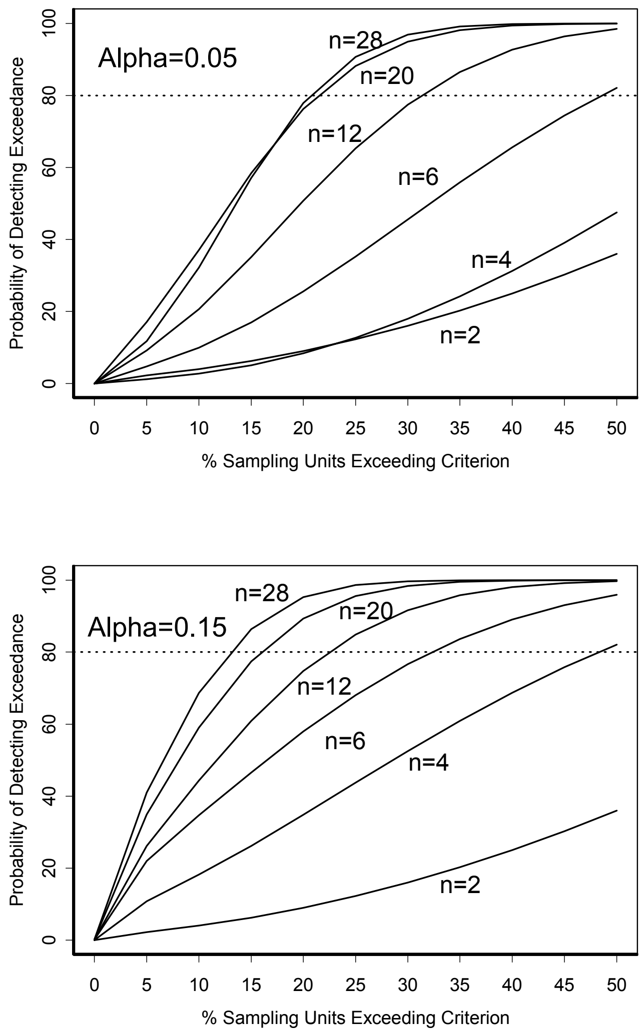

Figure 4-1. Power curves computed from different sample sizes, at two different alpha levels. In the upper panel, the Type I error rates (α) are held at ≤ 0.05, while those in the lower panel are held at $\alpha \leq 0.15$. The vertical axis is the probability of detecting that a sample exceeds the criterion exceedance rate, and the horizontal axis is the percent that a particular sample actually exceeds the criterion exceedance rate. Statisticians call the vertical and horizontal axes variables "power" and "detectable effect size," respectively. A horizontal reference line at 80% in each graph marks what many researchers consider the minimally acceptable power level for a statistical test. It is noteworthy that when $\alpha \leq 0.05$, none of the sample sizes shown has acceptable power for detecting exceedances $\leq 20\%$. However, when the acceptable Type I error is increased to $\alpha=0.15$, sample sizes of 28 and 20 have sufficient power for detecting exceedances between 13% and 20%. Smith et al. (2001) recommend that WQS tests should have sufficient power to detect exccedances of $\geq 15\%$ above criterion. The complex relationships among sample size, effect size, Type I error rates (α), and Power (1-Type II error rate) are discussed in detail in Appendices C and D.

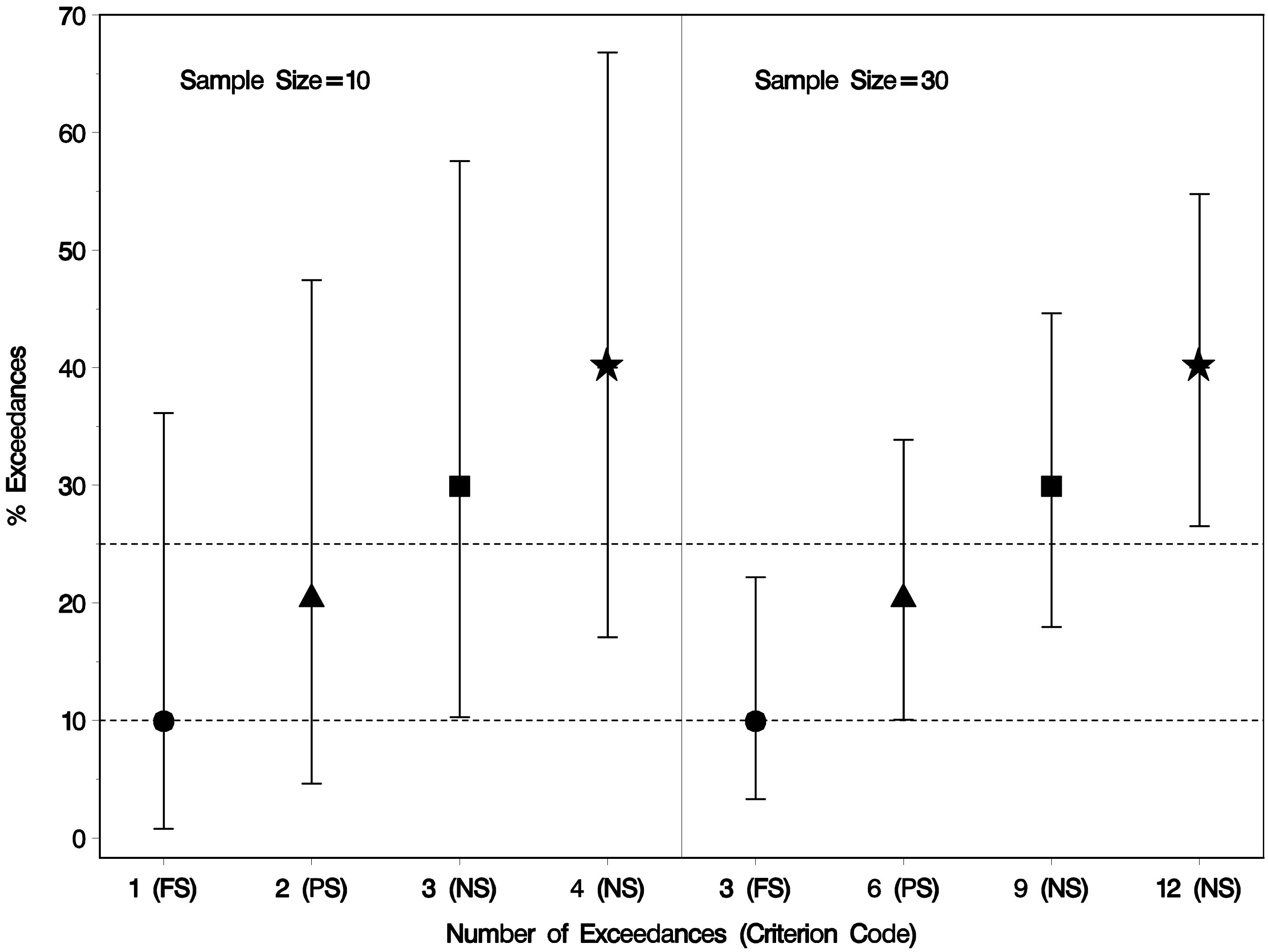

Figure 4-2. Effect of sample size on confidence interval widths. Two sets of two-sided 85% confidence intervals are illustrated (the 85% confidence level is recommended in Appendices C and D). For each sample size, confidence limits were estimated for samples with 10% (dot), 20% (triangle), 30% (square), and 40% (star) exceedances among their constituent sampling units. The two horizontal lines denote the 10% exceedance criterion and the 25% threshold (criterion value + 15% minimum detectable effect size) recommended by Smith et al. (2001). Exceedances of ≤10% constitute full support (FS) of the the 10% criterion; 10%-20% exceedances constitute partial support (PS); and exceedances >20% imply no support (NS). Any two-sided confidence interval whose lower bound extends below the 10% reference line satisfies the criterion. Thus with a sample size of 10, a minimum of 30% exceedances (= 3 sampling units) are required, compared with only 20% exceedances (= 6 sampling units) for a sample of 30 sampling units.

4.2.2 How Does the State Assess (Review and Evaluate) Data Quality?

The term "data quality assessment" means the scientific and statistical evaluation of data to determine whether data obtained from monitoring operations are of the right type, quality, and quantity to support water quality attainment decisions. Data quality does not exist in a vacuum; one must know in what context a data set is to be used in order to determine whether it is adequate.

Figure 4-1 shows the probability of detecting very low exceedances with small data sets, unless the actual rate of exceedance in the water is very high or common. Larger data sets have a greater probability of detecting less frequent exceedances. If a small data set detects an exceedance, the waterbody is likely experiencing a higher frequency of exceedances. However, if a small data set does not detect an exceedance, it is difficult to say with statistical confidence that the water is attaining WQS. Larger data sets are more powerful in terms of supporting decisions that a water is attaining WQS.

Guidance for assessing the quality of available data sets is provided in *Practical Methods for Data Quality Assessment* (EPA/600/R-96/084). For assessing WQS attainment, EPA recommends a tiered approach. The following steps should be part of the first tier of a data quality review process:

- Screen documentation to determine whether appropriate procedures were used and QA/QC measures were in place (e.g., if the third party's field and laboratory procedures are documented in SOPs).

- Determine whether samples were collected under the appropriate conditions for comparison with WQS (e.g., correct time of year or flow conditions).

- Review sample collection and analytical methods to determine compatibility with the state agency's QA/QC requirements and SOPs; also determine if the third party's sample collection and analytical methods were actually followed in creation of the data set.

- Determine whether the metadata accompanying the data set meet the agency's requirements (e.g., determine adequacy and accuracy of geographic documentation in the data set).

If the data do not meet all of the above conditions, they may be rejected from further analysis. Once it is determined that the data set meets the state's basic documentation requirements, the evaluators might decide to do additional screening of the actual data sets. At a minimum, they might want to look for values below the detection limit of the analytical method, because these may influence how the data set is analyzed or incorporated with other data. If, upon analyzing the data, evaluators suspect errors in the collection or analysis, they may want to conduct more in-depth analysis of QA/QC procedures. This screening could include reviews of QA/QC reports to determine if the data set meets the agency's QA/QC requirements regarding documenting measurement system performance (e.g., adequate use of QC samples), the approach to handling missing data and nondetects, and deviations from SOPs.

4.2.3 How Does the State Document the Level of Data Quality?

The 305(b) Consistency Workgroup developed a table assigning qualitative levels of information or data quality to different types of chemical data. Several states have since developed similar approaches for rating the quality of data used in WQS assessments. States are encouraged to use an approach similar to that described in Table 3-2 to report on the quality of data supporting attainment/impairment decisions. In addition, they should begin documenting quantitative information about the quality of these decisions.

The data hierarchy described in Table 4-2 addresses data quality considerations such as sample collection and analytical techniques, spatial and temporal representativeness, and QA procedures. The user rates the data set on the basis of the rigor of the information, where 1 is the lowest and 4 is the highest. In general, Level 1 information alone may not be sufficient for an attainment decision; however, even a short period of record can indicate impairment in cases of gross exceedances of criteria.

States should supplement the data descriptions illustrated in Table 4-2 with more quantitative descriptions of the confidence and power of their attainment/impairment decisions. This documentation clearly illustrates to decision makers and the public the impact of small data sets on the uncertainty in the water quality decision. Quantitative documentation of the uncertainty is expressed in statistical terms of error rates, both Type I decision error, or the α-level, and Type II decision error, or the β-level, of the assessment. These decision errors are discussed in detail in Appendices C and D. A Type I error occurs when an attaining waterbody is erroneously judged to be impaired, and a Type II error occurs when an impaired waterbody is erroneously judged to be attaining. EPA encourages states to collect sufficient numbers of samples to balance both types of error at reasonable levels.

To summarize, for attainment decisions based on chemical data, states should document:

- Level of information based on Table 4-2 or state-developed table or approach;

- Sample size, range of concentrations, mean, median, and standard deviation; and

- Level of statistical confidence (Type I decision error and Type II error) and width of the confidence interval.

4.3 How Does the State Analyze and Interpret Chemical Data to Determine WQS Attainment/Impairment?

The most important element of the state, territory, or authorized tribe's assessment and listing methodology is documentation of how the state analyzes data to determine WQS attainment and identify impaired waters. This documentation should be consistent with the state's, territory's, or authorized tribe's implementation procedures that are described either in the WQS or alternatively in other implementing regulations or policies and procedures documents. If the implementation procedures do not describe how WQS should be interpreted for determining

Table 4-2. Hierarchy of physical/chemical data levels for evaluation of aquatic life use attainment

Level of info[a]	Sample collection and analytical techniques	Spatial and temporal representativeness	Data quality
1	Any one of the following: • Water quality monitoring using grab water sampling • Water data extrapolated from an upstream or downstream station where homogeneous conditions are expected • Best professional judgment based on land use data, source locations	Low spatial and temporal coverage: • Quarterly or less frequent sampling with limited period of record (e.g., 1 day) • Limited data during key periods or at high or low flows (critical hydrological regimes)[b] • Data are >5 years old and are not reflective of current conditions	Approved QA/QC protocols not followed or QA/QC results inadequate Methods not documented Inadequate metadata
2	Any one of the following: • Water quality monitoring using grab water sampling • Rotating basin surveys involving multiple visits or automatic sampling • Synthesis of existing or historical information on fish contamination levels • Screening models based on loadings data (not calibrated or verified)	Moderate spatial and temporal coverage: • Bimonthly or quarterly sampling during key periods (e.g., spring/summer months) • Fish spawning seasons, including limited water quality data at high and low flows • Short period of record over a period of days or multiple visits during a year or season • Data are <5 years old and there is high certainty that conditions have not changed since sampling	Approved SOPs used for field and lab; limited training Low precision and sensitivity QA/QC protocols followed; QA/QC results adequate Adequate metadata
3	Any one of the following: • Composite or a series of grab water sampling used (diurnal coverage as appropriate) • Rotating basin surveys involving multiple visits or automatic sampling • Calibrated models (calibration data <5 years old)	Broad spatial and temporal (long-term, e.g., > 3 years) coverage of site with sufficient frequency and pollutant coverage to capture acute events: • Typically, monthly sampling during key periods (e.g., spring/ summer months, fish spawning seasons), multiple samples at high and low flows • Lengthy period of record (sampling over a period of months) • Data are <5 years old and there is high degree of certainty that conditions have not changed since sampling	Moderate precision and sensitivity Samplers well trained SOPs used for field and lab Moderate precision/ sensitivity QA/QC protocols followed; QA/QC results adequate Adequate metadata
4	Follows defined sampling plan which includes the following elements: • Description of how sample is representative of target populaiton • Defined data quality objectives, including error rate, confidence interval, sample size	Broad spatial (several sites) and temporal (long-term, e.g., > 3 years) coverage of site with sufficient frequency and coverage to capture acute events, chronic conditions, and all other potential chemical impacts • Monthly sampling during key periods (e.g., spring/summer months • Fish spawning seasons (including multiple samples at high and low flows) • Continuous monitoring • Data are <5 years old and there is high degree of certainty that conditions have not changed since sampling	High precision and sensitivity Samplers well trained SOPs used in field and lab QA/QC protocols followed; QA/QC results adequate Adequate metadata

[a]Level of information refers to rigor of chemical sampling and analysis, where 1 = lowest and 4 = highest.
[b]Even a short period of record can indicate a high confidence of *impairment* based on chemical data; 3 years of data are not required to demonstrate impairment. For example, a single visit to a stream with severe acid mine drainage impacts (high metals, low pH) can result in high confidence of impairment. However, long-term monitoring may be needed to establish full attainment.

attainment status, the procedures should be revised to, at a minimum, reference the assessment and listing methodology.

In recent years, most water quality agencies have followed approaches developed by the 305(b) Consistency Workgroup for interpreting data to assess WQS attainment/impairment status as described in the 305(b) reporting guidelines (U.S. EPA 1997). Guidance documents for developing section 303(d) lists of impaired waters indicate that waters identified as partially or not supporting WQS according to the 305(b) guidelines should be included on 303(d) lists of impaired waters (U.S. EPA 1991a). Under the Integrated Report, waters that are impaired or threatened for one or more designated uses can belong in Category 4 or Category 5 and may or may not require a TMDL depending on the source of impairment and the management action that has been completed on the waterbody. The Integrated Report addresses one area where the 305(b) and 303(d) guidances differed, the treatment of waters that are "fully supporting water quality standards, but threatened." The 305(b) guidance had a broader definition of waters fitting this category than 303(d) did, so it was not appropriate to assume that all threatened waters in 305(b) reports belonged on 303(d) lists. The Integrated Report clarifies there are only three instances in which waters that are threatened for one or more designated uses do not require the development of a TMDL: 1) if a TMDL has been completed, 2) if other pollution control requirements are reasonably expected to result in attainment in the near future, or 3) if the water is threatened by something other than a pollutant. In each case, follow-up monitoring should be scheduled for these waters to verify attainment of the WQS as expected. Table 4-3 reflects these modifications to the decision rules found in the 305(b) guidance made to simplify the reporting categories and to clarify the linkages between 303(d) lists of impaired waters and 305(b) water quality inventory reports. This table reflects the Integrated Report, therefore, does not include the "fully supporting, but threatened" category. The table also combines the "partially" and "not supporting" categories into a single category called "impaired."

An assessment methodology should take into account the balance between desired data requirements and the practical realities affecting the availability of information and the strength of the available evidence. For example, a state's methodology could describe an acceptable probability of decision errors for making an attainment decision, except in cases where overwhelming evidence of impairment is found. Examples of overwhelming evidence could be a single sampling event showing dangerously low pH downstream of an abandoned mine, fish kills that cannot be attributed to natural causes, elevated levels of accumulative pollutants in fish tissue. Another example could be allowing the results from analytical methods with high detection levels or poor sensitivity (e.g., field test kits) in cases where the results clearly suggest large exceedances of criteria. Photographs or other documentation of gross impairment may also be considered, if appropriate.

Generally, decisions should be based on very small data sets only when there is overwhelming evidence for impairment. EPA does not recommend making decisions based on small data sets of water column chemistry for attainment. Therefore, in the overwhelming majority of WQS scenarios, an approach based on probability sampling, in which states define an acceptable probability of decision error, will be preferred. Statistical inference based on sequential sampling designs may offer an alternative that allows states use defined data quality objectives to

Table 4-3. Interpreting chemical data to assess WQS attainment

Type of criteria	Attaining WQS	Impaired for 305(b) and 303(d)	Example of statistical guidelines for documenting data quality objectives for attainment decisions
Acute chemical criteria for toxic pollutant for the protection of aquatic life	For any one pollutant, no more than one excursion above acute criterion (EPA's criteria maximum concentration [CMC] or applicable state/tribal criterion) within a 3-year period based on grab or composite samples	More than one excursion above criterion within any 3-year period	One-sided binomial confidence intervals on the percentage of samples whose hourly mean exceeds the stated acute 1-hour mean criterion value. Type I and Type II error rates should be approximately equal to 0.15 and the minimum effect size set at 15% (0.15). The tests and confidence intervals evaluate: H_0: $\leq$5% of the samples exceed the 1-hour acute criterion value H_a: >5% of the samples exceed the 1-hour acute criterion value
Chronic chemical criteria for toxic pollutant for the protection of aquatic life	For any pollutant, no more than one excursion above chronic criterion (EPA's criteria continuous concentration [CCC] or applicable state/tribal criterion) within a 3-year period based on grab or composite samples.	More than one excursion above criterion within any 3-year period	One-sided binomial confidence intervals on the percentage of samples whose 4-day mean exceeds the stated chronic 4-day mean criterion value. Type I and Type II error rates should be approximately equal to 0.15 and the minimum effect size should be set at 15% (0.15). The tests and confidence intervals evaluate: H_0: $\leq$5% of the samples exceed the 4-day chronic criterion value H_a: >5% of the samples exceed the 4-day chronic criterion value.
Acute or chronic chemical criteria for conventional pollutant	For any pollutant, no more than 10% of the samples exceed the criterion	More than 10% of the samples exceed the criterion	One-sided binomial confidence intervals on the percentage of aliquots whose pollutant concentration exceeds the criterion value. Type I and Type II error should be approximately equal to 0.15 and the minimum effect size set at 15% (0.15). The tests and confidence intervals evaluate: H_0: $\leq$10% of the aliquots in the sample exceed the criterion value H_a: >10% of the aliquots in the sample exceed the criterion value.

Table 4-3. Interpreting chemical data to assess WQS attainment (continued)

Type of criteria	Attaining WQS	Impaired for 305(b) and 303(d)	Example of statistical guidelines for documenting data quality objectives for attainment decisions
Human health criteria for drinking water, fish consumption, recreation, or other human-health related uses	Annual mean concentration does not exceed criterion	Annual mean concentration exceeds criterion	Lower one-sided confidence intervals (or corresponding upper -one-sided t-tests) on the sample mean, geometric mean, or median pollutant concentration, pH, etc., relative to the stated criterion value. Type I and Type II error rates should be set at ≤0.15 and the minimum effect size at 15% (0.15). The tests and confidence intervals evaluate: H_0: mean, geometric mean, or median of the sample ≤ the criterion value H_a: mean, geometric mean, or median of the sample > the criterion value
Human health criteria for fish and shellfish consumption	Tissue levels do not exceed state/tribal risk-based levels, and/or water column concentrations do not exceed human health criteria	Tissue levels exceed state/tribal risk-based levels, and/or water column concentrations exceed human health criteria	Lower one-sided confidence intervals (or corresponding upper -one-sided t-tests) on the sample mean, geometric mean, or median pollutant concentration, pH, etc., relative to the stated criterion value. Type I and Type II error rates should be set at ≤0.15 and the minimum effect size at 15% (0.15). The tests and confidence intervals evaluate: H_0: mean, geometric mean, or median of the sample ≤ the criterion value H_a: mean, geometric mean, or median of the sample > the criterion value

identify impaired waters with small data sets. When a state describes its acceptable levels of decision error, it is able to identify the corresponding number of exceedances within a particular sample size that meet the level of decision error. With sequential sampling designs, the state, territory or authorized tribe may make an impairment decision once enough samples that fail to meet the WQS are collected and additional sample collection can be curtailed (see Figure 4-3).

Ecological applications of sequential sampling have been described (Carter et al. 1994); presumably this approach could be extended to many of the WQS situations described in this document.

4.3.1 What Statistical Analyses for Interpreting Chemical Data Does the State Use?

EPA acute and chronic chemical criteria for protection of aquatic life are examples of "ideal standards" as defined by Barnett and O'Hagan (1997). "Ideal standards" include criteria set as maximum levels not to be exceeded. As defined by Barnett, "ideal standards" pose several challenges in assessing attainment. The standard set as a not-to-be-exceeded chemical criterion does not address variation and uncertainty; therefore, assessing attainment implies a monitoring design that measures for the chemical throughout the entire population—all points in the waterbody continuously over time (Barnett 1997). Any state monitoring program to collect data for interpreting attainment with WQS, however, involves sampling the population and estimating the characteristics of the population on the basis of the characteristics of the sample. The use of sampling introduces variability and uncertainty. Some of this is due to the natural variability of the waterbody and human error associated with sample collection and analysis. A key element of the uncertainty relates to the precision of the sample. A larger, well-conducted monitoring effort will yield better, more precise estimates of the true condition than a smaller or poorly run effort. It is important that a state's data quality objectives and QA/QC procedures clearly define adequate statistical and other implementation procedures to ensure that all parties are aware of the minimum data set and statistical analysis requirements to show attainment (Barnett 1997). Figure 4-2 illustrates the effect of sample size on the confidence intervals and, therefore, the precision associated with attainment decisions.

Barnett recommends development of a statistically verifiable ideal standard composed of two parts. First is the ideal standard or criterion. Barnett provides several examples for defining the criterion based on a maximum, an average, or a percentile. The choice should reflect the pollutant–effect relationship. For example, pollutants that have a threshold effect should have a criterion that specifies that a high percentage of the samples must lie below the threshold in order to limit the amount of time the pollutant levels exceed the threshold (Barnett 1997).

The second component of the statistically verifiable ideal standard includes guidelines for statistical verification of the criterion (Barnett 1997). These guidelines, which should be a component of the state's WQS implementation procedures, describe a level of assurance that the criterion is attained. For example, the guidelines may describe the acceptable error rates for Type 1 and Type 2 error, the size of the confidence interval, and the sample size.

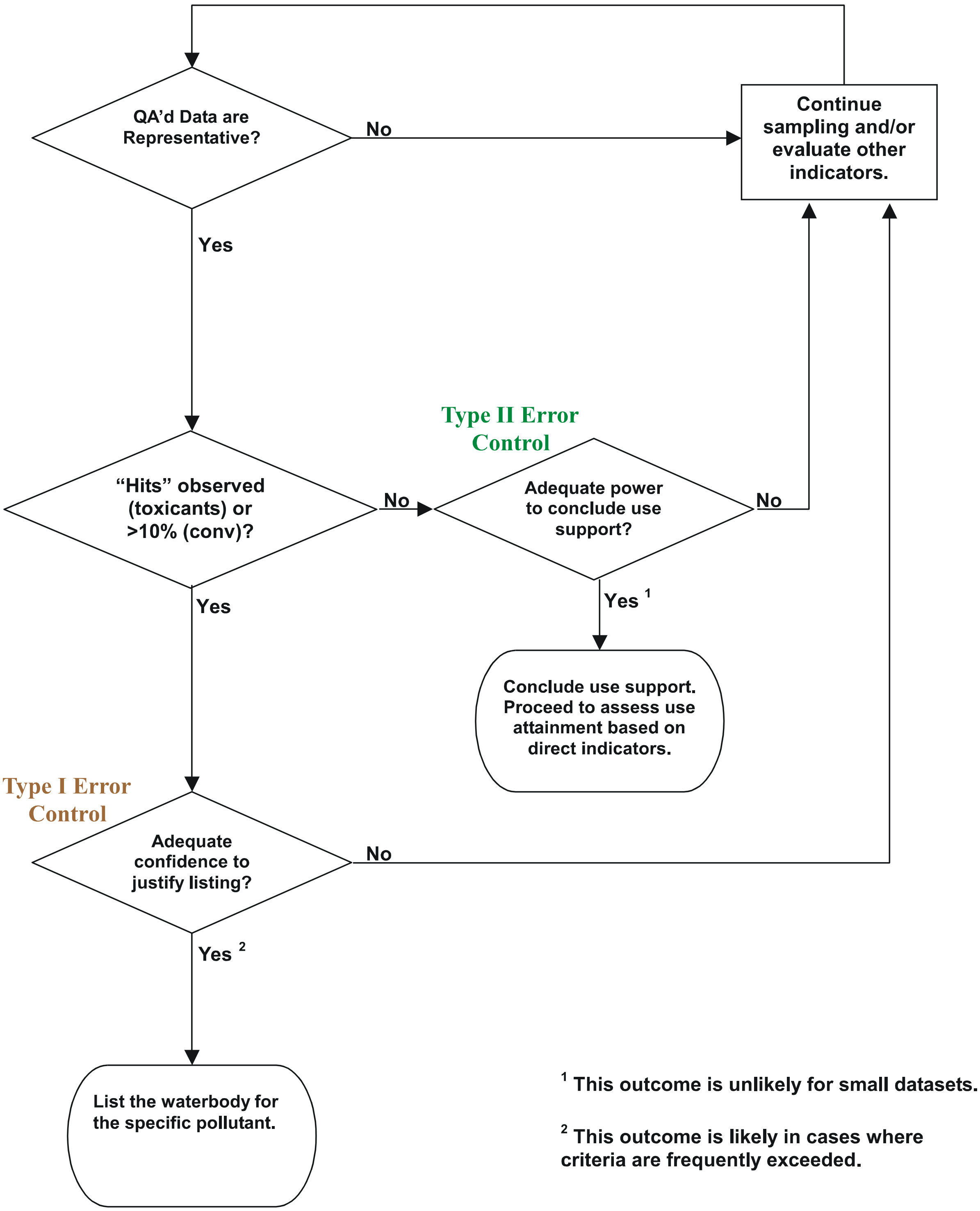

Figure 4-3. Sequential Decisionmaking. Making use support decisions, based on small sets of water column chemistry data, while balancing the risk of Type I (false positive) and Type II (false negative) decision errors.

Statistical methods are widely available to account for uncertainty and can be used to set appropriate bounds on how attainment should be demonstrated. Appendices C and D explore statistical hypothesis testing, confidence intervals, and Type I and Type II error and provides guidance on how to use parametric and nonparametric hypothesis tests to evaluate WQS attainment. The appendices include examples of the estimation of proportions of exceedances and upper concentration percentiles such as would be appropriate for comparison to threshold values. Methods for estimating mean and geometric mean concentrations and local variability and testing hypotheses about them are also included in the appendices. Tests and estimates of mean concentrations are appropriate for many of the human health criteria.

Another important analytical tool is trend analysis. Although state WQS generally call for analysis of data collected over 1- to 3-year periods in making attainment decisions, states should analyze data over longer time periods when they are available.

4.3.2 How Does the State Make Attainment/Impairment Decisions in the Absence of a "Perfect Data Set"?

State assessment and listing methodologies should describe how to make water quality attainment/impairment decisions in the absence of complete data sets that meet all their data quality requirements. For example, if a state's methodology calls for 30 samples to make an attainment decision with specified confidence, what will the state do if it collects only 10 samples? A state should develop procedures for looking for overwhelming evidence of water quality impairment, such as a single sample with well-documented QC methods that shows a large exceedance of an applicable criterion.

Another factor the state may want to consider is the effect of the sample size on the likelihood of detecting an exceedance of the criteria in the first place, particularly for chemicals that are not naturally occurring in the environment. For example, if a waterbody actually experiences 2 days during which a pollutant concentration exceeds the criterion, the probability that a sample of 36 daily samples collected over 3 years will capture both excursions is less than 1%. To detect such a small number of exceedances, one would have to collect about 600 samples. See Appendices C and D for further discussion and assumptions leading to the above statements about probabilities.

The state's methodology should also identify other types of data or information that may be used to support or supplement a sparse data set to make an attainment/impairment decision, such as the use of point-source discharge data from facilities discharge monitoring reports to predict ambient concentrations using dilution models. Regardless of the approach used, the state should clearly document how attainment decisions are made. If not documented elsewhere, the consolidated assessment and listing method is the appropriate place.

4.4 References

Barnett V, O'Hagan A. 1997. Setting environmental standards: The statistical approach to handling uncertainty and variation. London: Chapman & Hall.

Carter JL, Ravlin FW, Fleischer SW. 1994. Sequential egg mass sampling plans for gypsy moth (*Lepidoptera: Lymantriidae*) management in urban and suburban habitats. J Econ Entomol 87(4):999-1003.

Peterson SA, Urquhart NS, Welch EB. 1999. Sample representativeness: A must for reliable regional lake condition estimates. Environ Sci Technol 33:1559-1565.

Smith EP, Ye K, Hughes C, Shabman L. 2001. Statistical assessment of violations of water quality standards under Section 303(d) of the Clean Water Act. Environ Sci Technol 35:606-612.

U.S. Environmental Protection Agency (U.S. EPA). 1986. Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses. NTIS PB85-227049.

U.S. EPA. 1991a. Guidance for water quality-based decisions: The TMDL process. EPA/440/4-91-001.

U.S. EPA. 1991b. Technical support document for water quality-based toxics control. EPA/505/2-90-001.

U.S. EPA. 1994. Water quality standards handbook: 2nd ed. EPA-823-B-94-005A.

U.S. EPA. 1995. Final Water Quality Guidance for the Great Lakes System. Federal Register March 23, 1995; page 15365.

U.S. EPA. 1997. Guidelines for preparation of the comprehensive state water quality assessments (305(b) reports) and electronic updates. EPA-841-B-97-002A and 002B.

U.S. EPA. 1998. EPA guidance for quality assurance project plans. EPA/600/R-98/018.

U.S. EPA. 2000a. Guidance for data quality assessment. EPA/600/R-96/084.

U.S. EPA. 2000b. Guidance for the data quality objective process. EPA/600/R-96/055.

U.S. EPA. 2000c. Methodology for deriving ambient water quality criteria for the protection of human health. EPA-822-B-00-004.

U.S. EPA. 2000d. Use of fish and shellfish advisories and classifications in 303(d) and 305(b) listing decisions - WQSP-00-03. Transmittal memorandum of policy on use of fish and shellfish consumption advisories from Geoffrey Grubbs and Robert Wayland, October 24, 2000.