

US EPA ARCHIVE DOCUMENT

APPENDIX A: Detailed Review of the IWR Binomial Statistical Test

APPLICATION OF THE STATISTICAL TEST

A primary feature of the Florida Impaired Waters Rule (IWR) is the use of a statistical test based on the binomial distribution to evaluate data sets of water quality parameter measurements prior to relying on such data sets in listing a waterbody as “impaired.” Statistical tests are useful when making decisions based on limited information (samples) about a general condition (population). While samples generally represent a population, they may have limited power to accurately and precisely represent specific characteristics of that population with great confidence. For example, it can be difficult to determine whether a particular data set of water quality sample measurements accurately represents actual conditions in ambient waters.

The binomial distribution is a nonparametric test based on a yes/no or pass/fail outcome. Such tests can be used, for example, to determine how many defective parts are allowed to come off an assembly line run without rejecting the entire lot (the example given in Microsoft Excel software). Nonparametric tests are useful, in general, when data are sampled from a population that is not normally distributed (i.e., a “bell” shaped curve) or where some data are “off the scale” (i.e., too high or too low to measure because of limitations of measuring devices or detection limits). The latter condition is typical of many water quality data sets. Going back to the assembly line example, the binomial test as applied to water quality is used to determine how many “defective” water quality measurements can occur before the waterbody as a whole is determined to be impaired (rejection of the entire lot).

The binomial statistical test has two key components, a probability value and a confidence value (or alpha). The probability value represents the proportion of samples that do not meet applicable water quality criteria (or the proportion of “defective” samples) associated with determining impairment in the waterbody as a whole. In the IWR, the probability value is 10%. In other words, “I believe that a rate of 10% or more of samples not meeting water quality criteria is enough to determine that the waterbody as a whole is impaired”. The confidence value represents the desired certainty that small sample sizes are truly representative of the entire population. The confidence value is also expressed as a percentage value. In the IWR, the confidence value is 90% (80% for the planning list). In other words, “I want to be 90% certain that I have the right answer.” For small sample sets, application of the confidence value results in the proportion of samples not meeting criteria to be greater than 10% before determining impairment, because of the relatively low certainty that small sample sets adequately represent the waterbody as a whole. As the size of the sample set increases, the proportion of samples not meeting criteria that are necessary to determine impairment approaches 10% because of the increased certainty, afforded by more data, that the sample set adequately represents the waterbody as a whole. The choice of probability value is not affected by sample size: the same acceptable proportion of “defective” measurements is applied to large and small data sets. Likewise, the choice of confidence value is not related to the acceptable proportion of “defective” measurements: it is a separate expression of desired

certainty when considering the reliability of limited information. The probability value and the confidence value work together in the statistical test: “I want to be 90% sure that 10% or more of the samples do not meet water quality criteria in order to determine that the waterbody as a whole is impaired.”

INTERPRETATION OF THE PROBABILITY VALUE OF 10%

In 2005, EPA determined that changes to criteria were those that affected magnitude (i.e., “how much”; usually expressed as a concentration such as “milligrams per liter”), duration (i.e., “how long”; usually expressed as an averaging period in hours or days), and frequency (i.e., “how often”; usually expressed as a return interval such as “no more than once every three years” or as a percent of time), as these features establish the level of protection or underlying expectation for ambient water quality. EPA further determined that provisions related to data reliability or sufficiency were not changes to water quality standards. In 2005, and now, EPA has determined the confidence value is not a change to standards because it relates to data reliability rather than to magnitude, duration, or frequency. In 2005, however, EPA determined the probability value was a new or revised water quality standard as a change to the frequency component of criteria. As explained more fully below, EPA is changing that determination because, based on additional information submitted by FDEP, we believe the probability value is a data reliability component of the IWR rather than a modification to the frequency component of the criteria.

In evaluating the IWR, both the 2001 version examined in EPA’s 2005 Determination and the amended 2007 version which is the subject of this review, EPA’s question with respect to the binomial test is “what is meant by the probability value?”, or in other words, “what does it mean to be a ‘defective’ water quality measurement?” Is it defective in the sense that it is in error, inaccurate, biased, or an unreliable measure, or is it defective in the sense that it represents a pollutant or water quality parameter that exceeds its criterion? Based on the analytical framework laid out in EPA’s 2005 Determination, if it is the latter then the probability value represents a new or revised water quality standard as a frequency component of water quality criteria. Florida’s currently applicable water quality standards say that, “unless otherwise stated, all criteria express the maximum not be exceeded at any time.” However, if the probability value represents the former (data reliability), then it does not represent a new or revised water quality standard. Under this interpretation, the underlying expectations for the ambient water are unchanged: the criteria are not to be exceeded. The probability value establishes the strength of the signal from data that may include a proportion of unreliable measures that is necessary to conclude that the criteria have in fact been exceeded. In the absence of documented clarification, EPA acted expansively with respect to what is a new or revised standard and concluded that the probability value constituted a new or revised water quality standard in its review of the of the 2001 IWR (2005 Determination).

EPA now understands that the probability value operates differently than we determined it did in 2005. In 2005, EPA reasoned that application of the 10% probability value would result in a 10% exceedance of a criterion magnitude value in ambient water.

Under this earlier understanding, a “defective” measurement actually would represent a pollutant or water quality parameter that, in fact, would exceed the criterion in the ambient water. Requiring a 10% exceedance rate in the ambient water would be different than what is expressed in Florida’s water quality standards in terms of frequency. Based on consideration of additional information submitted by the State, however, EPA now understands that the purpose of the 10% probability value is to exclude data that are likely to be unrepresentative of actual ambient water conditions. Unless the number of samples ostensibly showing exceedance of the relevant water quality criterion is 10% or more, then FDEP will not list the receiving waters as having exceeded the criterion. The 10% probability value reflects the fact that the universe of samples assessed by FDEP are likely to include many unreliable and thus unrepresentative measurements, which do not accurately reflect the condition of the ambient water. Therefore, the State’s binomial statistical test requires 10% or more of such samples to exceed criterion magnitude values before it will determine the waterbody itself does not meet water quality standards.

MODIFICATIONS TO THE 2007 AMENDED IWR

The 2007 amended IWR differs from the 2001 IWR with respect to the binomial statistical test in both the wording of the rule language and the supporting rationale that the State submitted in 2007.

In the 2001 IWR, it was unclear whether the probability value component of the binomial statistical test revised the expectations for ambient water set out in Florida’s existing water quality standards. The binomial test provisions appeared in Florida Administrative Code (F.A.C.) rule 62-303.320(1), for the planning list, and rule 62-303.420(2), for the verified list, and the test was cross referenced in a number of other sections of the IWR.¹ The 2001 IWR described the probability value as “the number of exceedances of an applicable water quality criterion” necessary to determine impairment. EPA understood this language to revise the frequency component set out Florida’s existing water quality standards and, in its 2005 Determination, identified the provisions implementing the binomial as new or revised water quality standards.

The 2007 amended IWR addresses the binomial test in the same provisions of the Rule as did the 2001 IWR. However, the description of the probability value in the 2007 IWR refers to “the number of samples that do not meet an applicable water quality criterion” necessary to determine impairment for the waterbody as a whole. The consistent use of the term “samples” throughout these provisions describes the objective of the provisions as data reliability rather than ambient expectation. This interpretation is further clarified in the written materials submitted by FDEP in 2007.

The binomial statistical test first appears in the 2007 IWR in rule 62-303.320, related to the planning list. This provision has been renamed “Aquatic Life-Based Water Quality Assessment” in the 2007 IWR. The provision had been titled “Exceedances of

¹ Unless otherwise stated, all Rule and subsection citations are to provisions in the Florida Administrative Code.

Aquatic Life-Based Water Quality Criteria” in the 2001 Rule. The changes to the text in paragraph (1) are as follows:

Water segments shall be placed on the planning list if, using objective and credible data, as defined by the requirements specified in this section, the number of samples that do not meet exceedances of an applicable water quality criterion due to pollutant discharges is greater than or equal to the number listed in Table 1 for the given sample size. For sample sizes up to 500, waters are placed on the planning list when This table provides the number of exceedances that indicate a minimum of a 10% or more of the samples do not meet the applicable criteria exceedance frequency with a minimum of an 80% confidence level using a binomial distribution. For sample sizes greater than 500, the Department shall calculate the number of samples not meeting the criterion that are needed to list the waterbody with an 80% confidence level for the given sample size using the binomial distribution.

References to “number of exceedances” and “exceedance frequency” have been replaced with “number of samples”. Likewise, the changes in the text heading of Table 1 are as follows:

Minimum number of samples not meeting an applicable water quality criterion measured exceedances needed to put a water on the Planning list with at least 80% confidence that the actual exceedance rate is greater than or equal to ten percent.

The term “measured exceedances” and the phrase “that the actual exceedance rate is greater than or equal to ten percent” have been removed and replaced with “samples not meeting an applicable water quality criterion”.

The binomial statistical test appears in the 2007 IWR provisions related to the verified list at rule 62-303.420(2). This provision includes a 90% confidence limit, rather than the 80% confidence limit applied to the planning list. However, the probability value remains the same in this provision. Language changes similar to those made in rule 62-303.320(1) and Table 1 are also made for this provision and Table 3:

...Once these additional data are collected, the Department shall re-evaluate the data using the approach outlined in rule 62-303.320(1), F.A.C., but using Table 32, and place waters on the verified list when which provides the number of exceedances that indicate a minimum of a 10% or more of the samples do not meet the applicable criteria, exceedance frequency with a minimum of a 90% confidence level using a binomial distribution.

As with the changes to rule 62-303.320, the changes to rule 62-303.420 represent a clear change in meaning from the 2001 IWR. These changes in language clarify that the probability value of 10% is intended to be a data reliability provision related to the number of samples necessary to conclude that criteria have been exceeded in a waterbody

rather than a new allowable frequency of exceedance. EPA acknowledges that the assessment result is the same as in 2001. However, the amended language clarifies that the probability value of 10% serves as a data reliability provision related to the number of samples necessary to conclude that criteria have been exceeded in the waterbody as a whole rather than a new frequency component allowing ambient waters to exceed criteria 10% of the time. This clarification is fully explained in the FDEP supporting materials accompanying the submission of the IWR for review.

RELATED PROVISIONS IN THE 2007 AMENDED IWR

There are two important provisions within 62-303.320 that merit further discussion to understand the context of the application of the binomial statistical test. The first is paragraph (4)(a) which establishes a procedure for grouping data collected within a 4 day period and using the median as the representative value for the entire period. This provision clearly represents a new or revised water quality standard as it adds a duration component to the criteria. EPA reached the same conclusion in its 2005 Determination of the 2001 IWR, when the duration period was 7 days. The same duration period is established specifically for the marine dissolved oxygen daily average criterion in paragraph (5). The second note-worthy provision is paragraph (6)(b), which calls off the duration period in paragraph (4)(a) and the binomial statistical test for acute toxicity-based water criteria (as did the 2001 IWR) and for synthetic organic compounds and synthetic pesticides (which is new for the 2007 IWR), opting for a no more than once in three year period frequency of exceedance for any measurement above the criteria for any of these parameters. For practical purposes, these provisions limit the applicability of the binomial statistical test to metals, dissolved oxygen, and bacteria measurements.

Although they appear in planning list provisions, the duration and frequency criteria components described in 62-303.320(4)(a), (5), and (6)(b) constitute new or revised water quality standards based upon their cross reference in 62-303.420(1) and (6) and 62-303.720(m), which execute attainment decisions for purposes of meeting the requirements of Clean Water Act section 303(d).

The binomial statistical test described in 62-303.320, excluding the 4 day duration period, is cross referenced in 62-303.360(1)(a) and 62-303.370(1) for evaluating samples with respect to bacteria criteria and 62-303.380(1)(a) and (3)(a) with respect to drinking water and human health criteria (excluding synthetic organics and synthetic pesticides via 62-303.320(6)(b)). The binomial statistical test described in 62-303.420, excluding the 4 day duration period, is also cross referenced in 62-303.460(3)(a), 62-303.470(3)(a), and 62-303.480(3)(a) for evaluating samples with respect to bacteria criteria.

An important feature of the amended 2007 IWR is the so-called “overwhelming evidence clause” at 62-303.420(7):

...water segments shall also be included on the verified list if, based on representative data...scientifically credible and compelling information regarding

the magnitude, frequency, or duration of samples that do not meet an applicable water quality criterion provides overwhelming evidence of impairment.

This provision allows FDEP to consider data of known high quality and reliability, as well as data having other characteristics that make a credible and compelling case for non-attainment, and execute an attainment decision with respect to the 303(d) list. While this provision does not constitute a new or revised water quality standard, because the standards for evaluating the credible and compelling information are not changed, it does help provide needed flexibility for considering all relevant information pursuant to the regulatory requirements of 40 C.F.R. Part 130 for preparing an appropriate and complete list of impaired waters. There are also other provisions of the 2007 IWR that provide FDEP the legal authority to exercise discretion in identifying waters as impaired.

EVALUATION OF SUPPORTING RATIONALE

FDEP submitted a 40 page document entitled “Florida’s Methodology for Identifying Surface Water Impairment Due to Metals” (metals methodology) among the package of supporting material accompanying the submittal of the 2007 IWR for EPA review. In the Introduction section of this document, FDEP summarizes:

The IWR, which was adopted in 2001, establishes procedures for evaluating data sufficiency and data quality to ensure that a number of sample exceedances of a water quality criterion do, in fact, represent impairment of a waterbody. The statistical approach and thresholds selected are intended to provide greater confidence that the outcome of the water quality assessment is correct.

While the IWR uses EPA’s long-standing 10% exceedance rate as the threshold for impairment when evaluating aquatic life-based numeric water quality criteria, it differs from EPA’s Integrated Report guidance in two principal ways. First, it applies the threshold to both conventional pollutants and metals, while EPA recommends it only for conventionals. Florida applies this methodology to water quality parameters such as metals to account for uncertainty in data quality. Second, it establishes a minimum confidence level for the assessment (an 80% confidence level for the Planning List of potentially impaired waters and a 90% confidence level for the Verified List of impaired waters) that is calculated using a non-parametric statistical approach called the binomial method. (emphasis added)

Chapter 3 of FDEP’s metals methodology describes in detail the factors supporting the need to address uncertainty in data quality based on accounting for sampling and analytical error, with a particular concern for “false positive” (bias at the high end of measurement). The document states “erroneously high metal concentrations have routinely been reported in natural waters because of contamination artifacts introduced during sampling and analysis” (scientific literature citations provided). The document also states that “[i]t is the Department’s experience that much of the data reported for metals in natural waters are biased erroneously high and need to be verified

if reported to exceed water quality standards,” adding that “[s]ampling errors can sometimes be detected through metadata (for instance, if field blanks are contaminated).” Specific experiences related to working with Florida’s data set are recounted, as in:

The Department’s Bureau of Laboratories has referred a number of cases in which exceedances of water quality standards were alleged for metals; however further investigation (split sample studies, etc.) using analytical techniques designed to remove interfering substances (e.g., chelation extraction techniques for metals) nearly always demonstrated that measurement artifacts were the likely culprit, as few chronically reported water quality exceedances for metals could be substantiated in the laboratory or in properly designed field studies.

A detailed evaluation of phosphorus data from the Everglades provides some quantification of error rates from reports from lab analysis of field data, and the implications are summarized as:

While the previous example clearly illustrates the importance of metadata, the vast majority (>80%) of the state’s data providers still did not meet the metadata requirements of the original IWR due to data management constraints. FDEP has nonetheless accepted the data and has, in fact, revised the IWR to allow use of data without metadata because we do not want to overly limit the amount of data available for impaired water assessments². However, it should be noted that most of the water quality data collected for ambient waters come from laboratories with less incentive and less oversight than in the Everglades Program. Analysis of exceedances suggests that many are the result of data that were improperly qualified and that should not have been submitted without proper qualifiers identifying them as below the MDL or PQL. As a result, FDEP remains convinced that data lacking supporting QA/QC metadata (e.g., Legacy STORET data) should be used very cautiously in deciding whether a waterbody should be listed as impaired, and that the assessment methodology needs to acknowledge some level of false positives in the dataset. EPA’s TSD Response Summary states that **“the allowable frequency for criteria excursions should refer to true excursions of the criteria, not to spurious excursions caused by analytical variability or error.”**

When deciding on an appropriate assessment methodology, FDEP recognized that there would be some unknown number of false positives (given the potential for error combined with the limited ability to identify and exclude bad data). Because of the large water quality dataset (some 45 million records in the IWR database) it is not possible to do a QA analysis of each data point. As such, the only alternatives are to either exclude all data of unknown quality (the majority of currently available data), or to acknowledge this error in designing an assessment methodology. Florida’s methodology attempts to use as many data as possible to

² In cases where metadata show the data to be unreliable (i.e., do not meet the minimum QA/QC standards), the data are of course not used.

include as many waterbodies as possible in assessing waters for the TMDL Program. (emphasis theirs)

FDEP has assembled a large amount of data, a large proportion of which is from third party sources. This large database factors heavily in EPA's evaluation of the use of the binomial statistical test and FDEP's supporting material. Going back to the statistical background provided at the beginning of this analysis, the need for a method to determine the "greatest number of defective parts allowed to come off an assembly line run without rejecting the entire lot," or in this case "how many 'defective' water quality measurements need to occur to gain confidence that the water is impaired," is evident. FDEP's metals methodology provides an extensive list of outside data providers, along with the number of records provided by each. FDEP summarizes the challenges of working with large volumes of data from multiple sources:

Given the vast amount of ambient data available in Florida and the uncertainties associated with this data as far as its quality, accuracy and representativeness, FDEP needed to either limit the data that could be used to only that which could be rigorously evaluated for data quality and representativeness, or develop an assessment methodology that allowed for computerized, statistical evaluation of the data. Rather than limit the data that could be used, FDEP opted to use the vast combined monitoring capacity of multiple entities within Florida that collect data and promote documentation of collection, handling, and analysis, and reporting procedures.

However, from a practical management point, FDEP recognized that, even with improved sampling procedures, a significant fraction of the data will continue to represent erroneously high values because of errors introduced in sampling and analysis and bias from non-representative sampling. When examining data, it is not possible to identify (or program a computer to identify) which particular data points are valid or invalid because of the large range of possible results. However, certainty is increased greatly when multiple values are found to be exceeding a threshold. The extreme tail end of a distribution may be most likely to contain the most erroneous data, but as a greater proportion of the data lie above a threshold of interest, certainty increases greatly that the value has in fact been exceeded. The use of a 10% exceedance frequency in the IWR represents a threshold where the frequency of poor quality data suggests it is not likely that all the data above this point would be erroneously high, as a general rule. Thus, this serves as a practical adjustment for uncertainty from known data quality impacts, while ensuring confidence that waters that are impaired will be captured.

FDEP's methodology also documents and supports the selection of 10% as the probability value:

FDEP selected EPA's recommended 10% exceedance frequency as the listing threshold for the assessment of aquatic life use support in acknowledgement that some percentage of the available data are unreliable and/or represent natural

variation. The FDEP included the binomial method as a mechanism to establish the confidence associated with the assessment and applied the method to both conventional pollutants and toxics. FDEP has subsequently revised the IWR so that the binomial method does not apply to synthetic organics or pesticides because data for these pollutants are typically negatively biased. However, FDEP has concluded that the binomial method is appropriate for metals... The following points summarize FDEP's alternative approach for metals:

- The confidence limit aspect of the alternative approach using the binomial reflects FDEP's management of statistical uncertainty of sampling (grab sample monitoring) from an overall population (ambient water conditions)
- The 10% exceedance rate is a **sample** exceedance rate for the assessment data, not an inherent allowable rate of criteria exceedance in the ambient water. Florida must process over 45 million data records to conduct its assessment program, and nearly 75% of Florida's data are from other agencies. These non-FDEP data have greater uncertainty with respect to accuracy and representativeness, and it is not possible to thoroughly review the QA/QC associated with all these data. However, these data also provide a wealth of information about the status of Florida's waters. To most fully utilize these data resources, FDEP developed a statistical approach that is amenable to computerized data processing and that allows FDEP to achieve the objectives of using data most likely to be reliable, while ensuring that waters not expected to meet applicable water quality standards are indeed placed on the state's 303(d) list
- The 10% exceedance rate quantitatively represents an accounting for sampling and analytical error associated with factors such as collection and handling errors, reporting errors, blank contamination, reversals, and matrix interference. The extent and effect of these types of data quality factors have been quantified for specific data sets in Florida to provide further support for the selection of 10% as a reasonable and appropriate target value. For example, the USGS audit identified that 10% of the samples in Florida's data were unreliable. [*Note: this USGS audit was conducted using all of Florida's data, not just USGS collected data.*] The best quantification of potential error rates comes from Everglades data records, which indicate a range of between 2-60% for various water quality parameters. Excluding the extremes (a low overall error rate for calcium and a very high rate of blank contamination from one lab for orthophosphate), this range narrows to 7-33% with all but one remaining value above 10%. Recognizing that the majority of error is reflected on the high end of reported data, a selection of 10% is reasonable and appropriate for this accounting.

EPA finds this rationale reasonable and concludes that the 10% probability value does not constitute a new or revised water quality standard. EPA acknowledges that this conclusion differs from the 2005 Determination associated with the 2001 IWR with respect to the comparable provisions. However, EPA rigorously applied the identical analytical approach for evaluating what constitutes new or revised water quality standards as it employed in the 2005 Determination. With the benefit of FDEP's

supporting rationale and the changes in the regulatory language itself, the documentation of the 10% probability value functioning as a data reliability provision is clear and convincing. EPA believes that the characteristics of Florida's assessment data base in terms of volume of records and proportion generated from sources outside the state regulatory agency's control may be unique in the nation. While Florida has successfully made a State-specific case that use a 10% probability value in a statistical binomial test is appropriate and acceptable for use in Florida at this time, the documentation does not support this use as a general matter in other places or with an assessment data base that differs from Florida's current one in terms of documentation, quality, volume and underlying sources.

In its metals methodology, FDEP also makes an assertion concerning a minimal number of valid samples that exceed criteria, outside the context of data reliability:

The 10% exceedance rate also reflects that a minimal number of valid samples may exceed the criteria, but would not result in impairment of designated uses. No significant damage to the biological community is expected to occur from intermittent, low-level exceedances of chronic criteria because the exceedances are typically very short in duration (shorter than 96-hours) and, for metals, typically include non-bioavailable particulate forms. The results from FDEP stream bioassessments include many cases of waters that have had intermittent exceedances of chronic criteria for toxics and still have excellent bioassessment scores. Florida's well-developed bioassessment tools are an integral part of the assessment process, and FDEP believes that these tools are useful at identifying impairment of aquatic life use support.

This assertion no doubt expresses the belief of the authors of the report, but nonetheless does not have a relationship to the intended function of the 10% probability value, which is clearly identified as a "sample exceedance rate for the assessment data, not an inherent allowable rate of criteria exceedance in the ambient water" a few sentences above this assertion in the same Methodology document, nor did this assertion have any bearing on EPA's evaluation. However, as a factual matter EPA does not disagree with the general point, as evidenced by EPA's own criteria recommendation published pursuant to Clean Water Act section 304(a), which are the basis for the magnitude value in Florida's underlying water quality criteria for metals, and for which EPA has recommended associated duration and frequency components whereby the magnitude may be exceeded for short periods of time at infrequent intervals and still be fully protective of aquatic life uses. Florida could have elected to produce a methodology with an alternative allowable frequency component for their criteria, but they did not choose to do so.

CONTINUED EPA OVERSIGHT

While not identified as a new or revised water quality standard, EPA continues to have a responsibility for regulatory oversight of use of the 10% probability value in conjunction with its review of lists of impaired waters submitted to EPA pursuant to

Clean Water Act section 303(d). EPA recognizes that the 10% probability value represents a reasonable choice based on data quality as documented at this time. However, EPA also recognizes the improvement in data quality that Florida seeks in their underlying data moving forward, and that several provisions of the IWR encourage and mandate documentation of monitoring data used for water quality assessment purposes. EPA will continue to monitor and evaluate waters in all assessment categories with respect to the underlying data and the relevant aspects of the binomial statistical test as part of the Agency's oversight responsibilities under the Clean Water Act. EPA retains the discretionary authority to add waters to Florida's list of impaired waters if circumstances warrant. Furthermore, EPA will advise Florida accordingly if at some time in the future, continued use of the 10% probability value as a data reliability provision becomes inappropriate and counter-productive to Florida's program goals and responsibilities.

NATURALLY VARIABLE POLLUTANTS

As mentioned previously, the binomial statistical test applies to parameters other than metals, most notably to dissolved oxygen and bacteria criteria. EPA has addressed Florida's assessment methodology with respect to "naturally variable" pollutants or pollutant parameters in previous determinations and actions associated with Florida's 303(d) list. As explained above, EPA has determined that the binomial probability value is a "sample exceedance rate for the assessment data, not an inherent allowable rate of criteria exceedance in the ambient water." As to naturally variable parameters, like dissolved oxygen and bacteria, however, even if EPA determined the probability value were an allowable rate of criteria exceedance in a waterbody, that allowable exceedance rate would not constitute a new or revised water quality standard. As explained more fully below, applying a 10% exceedance rate to naturally variable parameters would be consistent with Florida's currently approved water quality standards and would not represent a change in magnitude, frequency, or duration.

Natural variability relates to the degree that conditions in nature vary as a function of time and space based on physical, chemical, biological, hydrological, and geomorphological factors. Pollutants and pollutant parameters can be placed into three distinct groups for considering the effects of natural variability. Some pollutants, such as chlorine and pesticides, are introduced solely as a function of anthropogenic activity and, although natural factors can mitigate or augment their effects, their presence cannot be attributed to natural conditions. The second group of pollutants usually occurs naturally in the environment at low levels, such as copper and cadmium, but protective water quality criteria for these pollutants usually lie well above the typical range of solely natural occurrence. For this group, the natural contribution is likely negligible at measured levels above or near the water quality criterion. Natural variability is generally not a factor for consideration in evaluating ambient measurement samples that exceed water quality criterion magnitude values for these first two groups of pollutants. By contrast, a third group of pollutants or pollutant parameters has protective water quality criteria that lie within or near the range of naturally occurring conditions. This naturally variable group includes pollutants or pollutant parameters such as dissolved oxygen,

turbidity, bacteria, conductivity, and alkalinity. Natural variability is an appropriate and reasonable factor to consider in evaluating ambient data for this group of pollutants or pollutant parameters.

Dissolved oxygen (DO) is perhaps the best example of a naturally variable pollutant parameter. DO refers to the volume of oxygen that is contained in water, and is measured and expressed as a concentration (typically in mg/L). Oxygen may occur in surface water as a by-product of photosynthesis by aquatic plants and/or through physical transfer from the surrounding air. DO solubility and, as a result, the expected ambient measured levels, are affected by temperature (colder water holds more oxygen), salinity (fresher water holds more oxygen), and altitude (lower pressure reduces oxygen's solubility). DO levels are also affected by flow and stream channel or lake morphology (more turbulent or well-mixed water transfers more oxygen from the air at the water surface), degree of biological activity (plant and animal respiration deplete oxygen, especially at night), and the amount of naturally occurring organic matter (aerobic decomposition depletes oxygen). As a result, DO can change and vary in a single water body according to time of day, season, weather, temperature, depth and location of sampling, and flow. The variability across different waters is augmented by many of the factors described above. DO can range from 0-18 mg/L in natural water systems, with long-term levels set generally within 5-6 mg/L to support a diverse aquatic community in most warmwater systems, as reflected by Florida's water quality standards.

An allowable exceedance rate of 10% for naturally variable pollutants would be consistent with EPA's general recommendations for such pollutants and would represent a reasonable choice for attainment decisions. In 2003, EPA approved, as consistent with Florida's existing water quality standards, FDEP's use of a 10% exceedance rate for naturally variable pollutants when compiling the State's Group 1 update to its section 303(d) list.³ The Eleventh Circuit Court of Appeals recently ruled in a challenge to that approval in *Sierra Club et al. v. Leavitt*, 488 F.3d 904 (11th Cir. 2007). One issue addressed by the Court was EPA's recognition that while some of Florida's water quality criteria are "not to be exceeded at any time," it was reasonable for Florida to interpret that regulatory phrase in concert with legislation authorizing the creation of Florida's water quality standards. That legislation provided that FDEP was to take into account the variability occurring in nature when applying the State's water quality standards. *Id.* at 919. The Eleventh Circuit held:

The EPA noted that because Florida does not have a monitoring program that continuously measures all points in its waterbodies (and thus the FDEP could never determine that a waterbody had not exceeded water quality criteria "at any time"), Florida must use statistical sampling to estimate a waterbody's compliance

³ See Decision Document Regarding Department of Environmental Protection's § 303(d) List Amendment Submitted on October 1, 2002 and Subsequently Amended on May 12, 2003. (June 11, 2003), page 25 and Appendix N on naturally variable pollutants.

www.epa.gov/region4/water/tmdl/florida/documents/EPA303d_decdoc.pdf

with water quality standards. Florida's Legislature recognized that sampling introduces variability into the testing process, some due to natural variability and some associated with sample collection and analysis. Thus, the EPA concluded, a single sample does not determine whether a waterbody fails to meet water quality standards. Instead, the EPA "considered a number of factors" in reviewing whether a waterbody was impaired. Decision Document at 21. "These factors included whether more recent data show attainment that renders earlier data suspect (trends); the magnitude of exceedance; the frequency of exceedance; pollutant levels during critical conditions; and any other site-specific data and information such as biological monitoring, whether new controls have been implemented on the water, etc." *Id.* Like the district court, we find the EPA's "totality" approach reasonable. *Id.* at 920. Recently, Florida has revised its underlying water quality standards to more clearly incorporate the legislative requirement that FDEP consider natural variability when applying its water quality standards:

In applying the water quality standards, the Department shall take into account the variability occurring in nature and shall recognize the statistical variability inherent in sampling and testing procedures. The Department's assessment methodology, set forth in Chapter 62-303, F.A.C., accounts for such natural and statistical variability when used to assess ambient waters pursuant to sections 305(b) and 303(d) of the Federal Clean Water Act. [Rule 62-302.530, F.A.C.]

EPA believes that Florida has correctly interpreted its own statute and regulations to recognize natural and statistical variability when making determinations of impairment. Therefore, even if EPA were to determine that the 10% probability value in the binomial statistical test was a new allowable exceedance rate rather than a data reliability provision, EPA would also determine such an exceedance rate does not constitute a new or revised water quality standard as to naturally variable pollutants.

Bacteria represents a special case in applying the binomial statistical test because the criteria itself includes allowable exceedance rate of 10% in ambient water. In this case, application of the 10% probability value is redundant with the criteria already in place as a practical matter. It is clear there is no intended change in criteria. EPA considers the application of the 10% probability value to provide no additional consideration for data reliability as a listing methodology for this component of the bacteria criteria. The binomial statistical test does function to add a confidence value to the assessment procedure. Regardless, however, EPA is neither approving nor disapproving the confidence value because it is not a not a new or revised water quality standard.

USE OF THE CONFIDENCE VALUE

As described in the beginning of this appendix, the confidence value represents the desired certainty that small sample sizes are truly representative of the entire population. In a few places in its 2005 Determination, EPA mistakenly suggested that the

application of the confidence value constituted a new or revised water quality standard. For example, on page 14 of Appendix C of the 2005 Determination, EPA stated:

EPA has determined that as applied to Shellfish Use Consumption Support, this provision changes or further defines the frequency of Florida's currently approved Fecal and Total Coliform criteria found at 62-302.530(6) and (7) from a strict "not more than 10% of the samples exceeding . . ." and replaces it with an evaluation of samples targeting higher than 10% of the samples to gain confidence of an actual exceedance rate of 10%.

On pages 55-56 of that same document, EPA stated:

EPA does not find the minimum sample size aspect of this provision to be a water quality standard. This provision relates to the exclusion of data for CWA 303(d) listing purposes pursuant to implementing regulations at 40 CFR Part 130.7(b)(5) and 40 CFR Part 130.7(b)(6)(ii) and (iii). This aspect of the provision is not a water quality standard because it does not describe the ambient condition of a water body. This provision contains policy choices about what data is reliable, but it does not describe the condition of the water body that is assessed. Additionally, applying a confidence test to assessing exceedance frequency does not itself change the targeted magnitude, duration, and frequency of criteria that describes the ambient condition of the waterbody as long as the targeted exceedance frequency is equivalent to the underlying frequency of the existing water quality standard. The statistical confidence test relates to the reliability or sufficiency of data rather than to the ambient condition of the waterbody. The statistical confidence takes into account the variability of data that derives from sampling error that occurs in any field sampling/water monitoring, and thus whether the data accurately represent the condition of the waterbody, but it does not incorporate a different ambient condition in the waterbody - in other words, a different level of pollutant(s) or pollutant indicators that are acceptable in the waterbody. The frequency of exceedance, however, does relate to the ambient condition and therefore is a part of a water quality criterion. The statistical confidence test may be used to gain assurances of an exceedance of a defined frequency for purposes of identifying water quality limited segments. [emphasis added]

The underlined portion of the second quote above reflects the correct understanding of the confidence value and EPA's current determination with respect to whether the confidence value constitutes a new or revised water quality standard. However, the rationale offered in the next sentence of the 2005 Determination, "statistical confidence takes into account the variability of data that derives from sampling error that occurs in any field sampling/water monitoring, and thus whether the data accurately represent the condition of the waterbody," does not correctly describe how the confidence value works in the IWR. A statistical confidence test does not account for the underlying accuracy of data, rather it accounts for the representativeness of the sample data -- how well a sample

set represents a population. The effect of sampling error is accounted for by the probability value in the IWR.

As explained above, FDEP demonstrated that 10% is a reasonable representation of erroneously high values in their overall population of water quality data, without respect to sample size. If one could expect 10% of the data to be in error regardless of sample size (i.e., a 10% error rate for the population of recorded ambient measurements), then a confidence value associated with sample size simply represents the degree to which a small sample set could disproportionately represent erroneously high values (i.e., the sample set may have more than 10% erroneously high values while the population maintains an overall rate of 10% erroneously high values). Thus, the confidence value component of the binomial statistical test does not constitute a new or revised water quality standard in any context that it appears in the IWR.