85

A Nonparametric Procedure for Listing and Delisting Impaired Waters Based on Criterion Exceedances

Pi-Erh Lin, Duane Meeter, and Xu-Feng Niu

Department of Statistics
Florida State University
Tallahassee, FL 32306-4330

October 2000

Technical Report Submitted to
The Florida Department of Environmental Protection
for the Fulfillment of Task 1, Contract No. LAB015

A Nonparametric Procedure for Listing and Delisting Impaired Waters Based on Criterion Exceedances

Pi-Erh Lin, Duane Meeter, and Xu-Feng Niu

October 2000

1. Purpose

This technical report provides the Florida Department of Environmental Protection (Department) with a description of the theoretical foundation for its proposed statistical methodology for determining impairment based on water quality criterion exceedances. A similar description for the identification of waters that are no longer impaired is also provided. Based on statistical analysis, it is recommended that a minimum of ten samples be required for listing an impaired water body and that a minimum of 28 samples be required for delisting. Using these recommended minimum samples, the listing and delisting decisions are correct with approximately 95% level of confidence.

2. Background Information

Section 305(b) of the Clean Water Act (CWA) requires states to conduct water quality surveys to determine whether or not their water bodies are healthy and of sufficient quality to meet their designated uses. The United States Environmental Protection Agency (USEPA) collects and utilizes this information to prepare a biennial report, known as the National Water Quality Inventory (or more commonly referred to as the "305(b) Report"), for the Congress of the United States.

Section 303(d) of the CWA requires states to prepare lists of "surface waters that do not meet applicable water quality standards", referred to as **impaired waters**, and to establish Total Maximum Daily Loads (TMDLs) for pollutants causing the impairment of these waters on a prioritized schedule. A TMDL establishes the maximum daily amount of a pollutant that a water body can assimilate from all sources without causing exceedances of water quality standards. As such, the development of TMDLs is an important step toward restoring surface waters to their designated uses.

The 1999 Florida Watershed Restoration Act clarified the Department's authority for the TMDL program and directed the Department to develop a methodology, and adopt it by rule, that clearly defines those waters that should be included in the state's 303(d) list of impaired waters. Given the importance of the TMDL program, the Department formed a Technical Advisory Committee (TAC) for the purpose of developing a clear, consensus-based method to define impaired lakes, streams, and estuaries. Members of the TAC were selected based on their technical expertise in key scientific fields. While the resultant 303(d) list will directly determine which waters are to be targeted for TMDL development, the list could be used to help prioritize a variety of other watershed restoration efforts in Florida.

One important measure of water body health is the concentration of conventional pollutants, metals, and dissolved oxygen. Conventional pollutants include chlorides, total fecal coliform, and fluoride. Metals include arsenic, aluminum, cadmium, chromium, copper, iron, lead, mercury, nickel, selenium, silver, thallium, and zinc. Florida's surface water quality criteria are used to assess whether a pollutant or a metal level is too high (or too low for the case of dissolved oxygen) to preclude the water body from meeting its designated uses. More specifically, a state regulatory agency may wish to set a water quality criterion for each pollutant and each metal, and refer to a single observation or measurement as an ***exceedance*** if it exceeds the criterion.

Based on guidance provided by the USEPA, which recommends a "greater than 10% exceedance percentage" for determining that waters only partially meet their designated use for aquatic life use support, the TAC developed a methodology for the listing and delisting of impaired water bodies depending on whether or not the true exceedance percentage is larger than 10%. However, the true exceedance percentage of a pollutant or metal in a water body reach is usually unknown, and must be estimated from random samples. The key question raised by the TAC was *"How do we draw a highly reliable statistical conclusion on the true exceedance percentage based on sample exceedance percentage?"*

The current study will address this and related issues based on statistical methods. In the study, the words "chance", "percentage", "probability", and "proportion" will be used interchangeably. They are used to describe the likelihood of an event and are related in the following way:

$$\text{Chance} = \text{Percentage} = (\text{Probability})\times 100\% = (\text{Proportion})\times 100\%, \qquad (2.1)$$

where Probability = Proportion is expressed as a real number between 0 and 1. For example, the probability (or proportion) of raining today is 0.7 but the chance (or percentage) is 70%.

The Florida 305(b) Report is prepared using the STORET water quality database, and biological data from the state's biology and rapid bioassessment sampling programs. It should be noted that the available data sets for key water quality parameters are quite small for many Florida water bodies over a five-year period. For example, over the five-year period from 1994 to 1998, 590 out of 849 (69%) water reaches had organic nitrogen sample sizes ranging from 1 to 20, and 568 out of 983 (58%) water bodies had dissolved oxygen sample sizes ranging from 1 to 20. Detailed information on available sample sizes is listed in Table 1 for six pollutants: organic nitrogen, dissolved oxygen (DO), Ammonia (NH_4), total nitrogen, total phosphorus, and nitrate (NO_3). Given these small sample sizes, any proposed listing and delisting procedures, based on the calculated sample excceedance percentages, must be applicable to both large and small samples.

For a given pollutant or metal in a water body, the sample proportion of exceedances is a point estimator of the ***true exceedance probability p*** for the pollutant or metal. Since the estimator varies in a random manner from sample to sample, inferences about the true exceedance probability based on the estimator will be subjected to uncertainty. The degree of uncertainty depends on the exceedances and the sample size: the smaller the sample size is, the greater the uncertainty will be. Therefore, the sample proportion of exceedances should not be used for the determination of water body health without considering its sample size. The reliability of the estimated exceedance probability relating to sample size should be addressed.

In this study, a nonparametric procedure is proposed for listing and delisting impaired water bodies based on criterion exceedances and sample sizes. The uncertainty of estimated exceedance probabilities is examined, and tests of hypotheses about the true exceedance probabilities of pollutants and metals are performed. The proposed nonparametric procedure provides a scientific approach for identifying impaired surface waters based on the measured percentage of exceedances of water quality criteria. Specifically, in Section 3, a nonparametric procedure for listing impaired waters is proposed using both a confidence interval approach and a test of hypothesis approach. A nonparametric procedure for delisting is proposed and discussed in Section 4. The delisting procedure is not a mirror image of the listing procedure because a much larger sample size is required for delisting than for listing impaired waters at a comparable level of confidence. Concluding remarks and discussion are provided in Section 5. The proposed nonparametric listing and delisting procedures are equally applicable to both conventional and toxic pollutants.

3. Listing Procedure

The TAC recommended that a water body reach be listed as impaired whenever the true exceedance probability of a pollutant or metal is greater than 0.1. This recommendation will be referred to as "the 10%-exceedance method." With respect to a criterion threshold, a single observation of a pollutant takes one of two values: "**yes**, the measurement exceeds the threshold" or "**no**, it does not". Of course, the actual distribution of a pollutant measurement in a water body is usually unknown. However, using the number of measured exceedances, the unknown distribution of a pollutant measurement can be transformed to a ***binomial distribution*** that depends only on the sample size and the true exceedance probability p. For example, a single observation for copper can take one of two values: "**yes**, the measurement exceeds the copper threshold of 2.9 $\mu g/l$" or "**no**, it does not". An important question arises for the regulatory agency. That is, how many exceedances out of n samples indicate the water exceeds the true exceedance percentage (e.g., 10%) that has been established to constitute impairment of the designated use? Note that deciding whether or not a single observation is a criterion exceedance is a different thought process than determining the minimum number of exceedances for determining impairment. In developing a listing procedure, the following two approaches were considered.

a. Confidence Interval Approach

In general, a ***binomial distribution*** is defined for experiments that result in a dichotomous response, i.e., responses for which there exist two possible alternatives, such as yes-no or pass-fail. A binomial random variable, X, which represents the total number of **yes** responses, has the following characteristics: (1) the experiment consists of n identical trials, (2) the trials are independent, and (3) the probability of **yes** remains the same from trial to trial. In this study, a "trial" refers to a single sample taken from a water body reach and the probability of **yes** response for a single trial is denoted by p, which is also the true exceedance probability of a pollutant. Thus, the probability of **no** is $1-p$. For a binomial random variable X with n trials, the mean (or expected value) and variance of this variable are np and $np(1-p)$, respectively. The square root of the variance, $\sqrt{np(1-p)}$, is called the standard deviation of the binomial random variable. Both variance and standard deviation measure the variability of a given random variable. For a particular water body reach, the probability p of an observed pollutant exceeding its criterion threshold depends on the unknown distribution of the pollutant and must be estimated. It is well-known that the sample proportion of **yes**, denoted by $\hat{p}$ = (total number of **yes** responses)/(sample size) = X/n, is the best point estimator for the true exceedance probability with expected value p and standard deviation $\sqrt{p(1-p)/n}$. The estimator is "best" in the sense that it is unbiased and has the minimum variance among all unbiased estimators. However, the estimator $\hat{p}$ itself is a random variable varying from sample to sample. Using it for the estimation of p often results in a "hit and miss" scenario and is not reliable. Modern statistics strongly recommends the use of a confidence interval estimation approach that takes into account the variability of the estimator.

The most commonly used interval estimator is a two-sided confidence interval. But, in environmental or ecological applications, it is often more cost-effective to obtain a one-sided confidence interval to assess the true exceedance probability p for the compliance of regulations. Both the two-sided and one-sided confidence intervals are described below. However, in this study, attention will be focused on the one-sided intervals for listing and delisting impaired water bodies.

Two-Sided Confidence Interval: Let [L, U] denote a two-sided $(1-\alpha)100\%$ (e.g., 95%) confidence interval for p where L and U are the lower and upper limits, $0 \leq \mathrm{L} \leq \mathrm{U} \leq 1$, and α is a significance coefficient satisfying the following probability inequality,

$$P(\mathrm{L} \leq p \leq \mathrm{U} \mid n, X) \geq 1 - \alpha, \tag{3.1}$$

with the interval length, U – L, being the shortest when the number of exceedances is observed. Note that both L and U depend on the sample size and the number of exceedances, X, and hence are random variables. The probability inequality in (3.1) is used since X is an integer random variable and the prescribed probability of $(1 - \alpha)$ may not be reached exactly by any integer observation.

One-Sided Confidence Interval: There are two types of one-sided confidence intervals that can be constructed; a **lower** one-sided $(1-\alpha)100\%$ (e.g., 95%) confidence interval for p is given by [0, U] and an **upper** one-sided $(1-\alpha)100\%$ confidence interval for p is given by [L, 1] where L and U can be computed as follows. Let x denote the observed number of exceedances in a water body. Then,

$$\mathrm{L} = \text{largest } p \text{ such that } P(X \geq x \mid n, p) \leq \alpha,$$

or

$$\mathrm{L} = \text{largest } p \text{ satisfying } P(X \leq x-1 \mid n, p) \geq 1 - \alpha, \tag{3.2}$$

and

$$U = \text{smallest } p \text{ such that } P(X \le x \mid n, p) \le \alpha. \tag{3.3}$$

Two-sided confidence intervals for an exceedance probability p can be obtained following the procedure of Blyth and Still (1983), where a table of two-sided 95% and 99% confidence intervals is provided for $1 \le n \le 30$. Using the table, when the sample size is $n = 10$ and the number of exceedances is $x = 3$, i.e., $\hat{p} = 0.3$, the two-sided 95% confidence interval for p is found to be [0.09, 0.62] and the two-sided 99% confidence interval is [0.05, 0.70]. These intervals can be obtained by an application of (3.2) and (3.3) with α replaced by $\alpha/2$. Under the same example, when $n = 10$, and $x = 3$, an upper one-sided 95% confidence interval [L, 1] for p can be obtained by the use of (3.2) as follows: From (3.2), the lower limit L is calculated as the largest p satisfying the probability inequality,

$$P(X \ge 3 \mid n, p) \le 0.05 \quad \text{or} \quad P(X \le 2 \mid n, p) \ge 0.95. \tag{3.4}$$

Using a computer program, e.g., MINITAB, for binomial distribution, with $n = 10$ and $x = 3$, the value of L is found to be 0.08725. Thus, the upper one-sided 95% confidence interval for p is [0.08725, 1.0], i.e., $p \ge 0.08725$. (If a binomial probability table is used, an interpolation method may be required. With binomial probabilities listed for $p = 0.05$ and 0.1, an approximate value of L is found to be 0.0828.)

It is important to understand the meaning of a confidence interval since it is often misunderstood and incorrectly interpreted in practice. In particular, it is important not to use the word "confidence" as a synonym for the word "chance" or the word "probability". Referring to the above example, it is <u>not correct</u> to say that there is a 95% chance that the true exceedance probability p will fall in between 0.0827 and 1.0." A correct interpretation is that we are 95% confident that the true exceedance probability p falls in the interval [0.0827, 1.0]. The "95% confidence" refers to the fact that, in repeated sampling, approximately 95% of all similarly constructed intervals will enclose the true exceedance probability, p. The remaining 5% will not." Suppose that, for the sake of explanation, there is available a total of 1000 random samples each of size $n = 10$. Using the same probability inequality (3.4), 1000 intervals can be constructed. Of these, about 950 (= 1000 $\times$ 0.95) intervals will enclose the true exceedance probability p (call these "good intervals"), but the remaining intervals will not. If we randomly select one sample of size $n = 10$ resulting in the interval [0.0827, 1.0], the odds are 19 to 1 (simplified from the odds of 0.95 to 0.05) in our favor that we have selected one of the roughly 950 "good intervals." In other words, the probability is 0.95 that the constructed interval [0.0827, 1.0] is from the pool of about 950 "good intervals". While we do not have 100% certainty that the interval [0.0827, 1.0] includes the true exceedance probability p, we are 95% **confident** that the interval [0.0827, 1.0] **does include** p. In this case, we conclude that, with 95% confidence, $0.0827 \le p \le 1.0$, definitely!

Note that, for an upper one-sided confidence interval, and for a fixed n and given x, the value of p and the $(1-\alpha)100\%$ level of confidence are related by the following inequality,

$$P(X \le x-1 \mid n, p) \ge 1 - \alpha. \tag{3.5}$$

Using (3.5) with $p = 0.10000001$ (to mean $p > 0.1$), $n = 10$, and $x = 3$, it follows that $1-\alpha = 0.9298$. That is, an upper one-sided 92.98% confidence interval for p is [0.10000001, 1.0], or $p > 0.1$. The above illustration shows that, if three or more exceedances are observed among 10 measurements, then with approximately 93% confidence, the water body will be listed as impaired using the 10%-exceedance method. In the current study, the minimum numbers of exceedances, x, required for the lsting of a water body reach as impaired, with approximately 95% confidence, are proposed for various sample sizes n, $1 \leq n \leq 100$. They are given in Table 2. It should be noted that the actual confidence level is not 95% because we are rounding off to the nearest whole number of exceedances and that the confidence level varies from sample size to sample size.

This confidence interval approach could be adopted to develop a set of guidelines for the listing of impaired waters as demonstrated above. A second approach is based on the test of hypothesis.

b. Test of Hypothesis Approach

Testing a hypothesis about exceedance probability is an alternative way to assess an estimator and its uncertainty. Suppose that, for a particular pollutant, two out of ten measurements in a water body exceed the criterion threshold. Is the sample exceedance percentage of 20% (i.e., $\hat{p} = 0.2$) strong evidence to determine the water body as impaired using the 10%-exceedance definition of impairment? Or, equivalently, is the sample percentage of 20% significantly larger than an assumed true exceedance percentage of 10% based on only $n = 10$ measurements? This question can be put in the framework of hypothesis testing. Here, we wish to test the null hypothesis

$$H_0: p \leq 0.1, \qquad (3.6)$$

that is, the water body is not impaired, versus the alternative hypothesis

$$H_a: p > 0.1, \qquad (3.7)$$

that is, the water body is impaired. The test can be performed by referring the observed number of exceedances, x, to a binomial probability table. When $n = 10$ and $p = 0.1$, the probability of observing two or less exceedances is 0.9298 (and the probability of observing three or more exceedances is 0.0702). If the number of exceedances in the ten measurements is two or less, the sample does not provide sufficient evidence to reject the null hypothesis. Thus, the sample 20% is not significantly larger than the assumed 10% exceedance percentage. But, if three or more exceedances are observed, there is sufficient evidence to conclude that, at the 7% significance level, the true exceedance probability p in the water body reach is over 0.1, and the alternative hypothesis H_a: $p > 0.1$ is accepted. That is, a 30% sample exceedance percentage is significantly larger than the assumed 10% exceedance percentage at the 7% level of significance. This is equivalent to saying that a 93% confidence interval would exclude $p \leq 0.1$ when there are three exceedances in a sample of ten.

As mentioned in the beginning of Section 3, the TAC recommended a 10%-exceedance definition of impairment. If the recommendation is adopted into the rule, the Department will need to provide ***a set of guidelines on the minimum number of exceedances and sample size required for listing impaired waters.*** In the above example, when ten samples are collected from a water body and analyzed, the minimum number of exceedances required to list the water body as impaired is $x = 3$, with approximately 93% confidence. Using the test of hypothesis approach, the water body will be listed as impaired whenever the random sample results in the acceptance of H_a: $p > 0.1$ at a suitable $100\alpha\%$ significance level or, equivalently, at a suitable $(1-\alpha)100\%$ confidence level. The results turn out to be identical to those obtained by the use of the confidence interval approach. The fact that the two approaches produced identical results for listing impaired water bodies, as presented in Table 2, is not surprising. It is due to the duality relationship between the confidence interval and test of hypothesis approaches. See, e.g., Bickel and Doksum (2001, Section 4.5) for a detail explanation of the duality.

It should be noted that the minimum numbers of exceedances for listing an impaired water body given in Table 2 can be generated by many statistical packages. The Microsoft Excel function CRITBINOM(trials, probability_s, alpha) calculates the smallest number of successes "k" out of "n" trials when the probability of a **yes** response on each trial is p such that $P(X \le k \mid n, p) \ge$ alpha. Here, "k" and "alpha" are, respectively, equal to "$x - 1$" and "$1 - \alpha$" of (3.5). The CRITBINOM(n, p_0, $1-\alpha$) function provides the critical value, $x = k + 1$, for the test of null hypothesis

$$H_0\colon\ p \le p_0, \tag{3.8}$$

versus

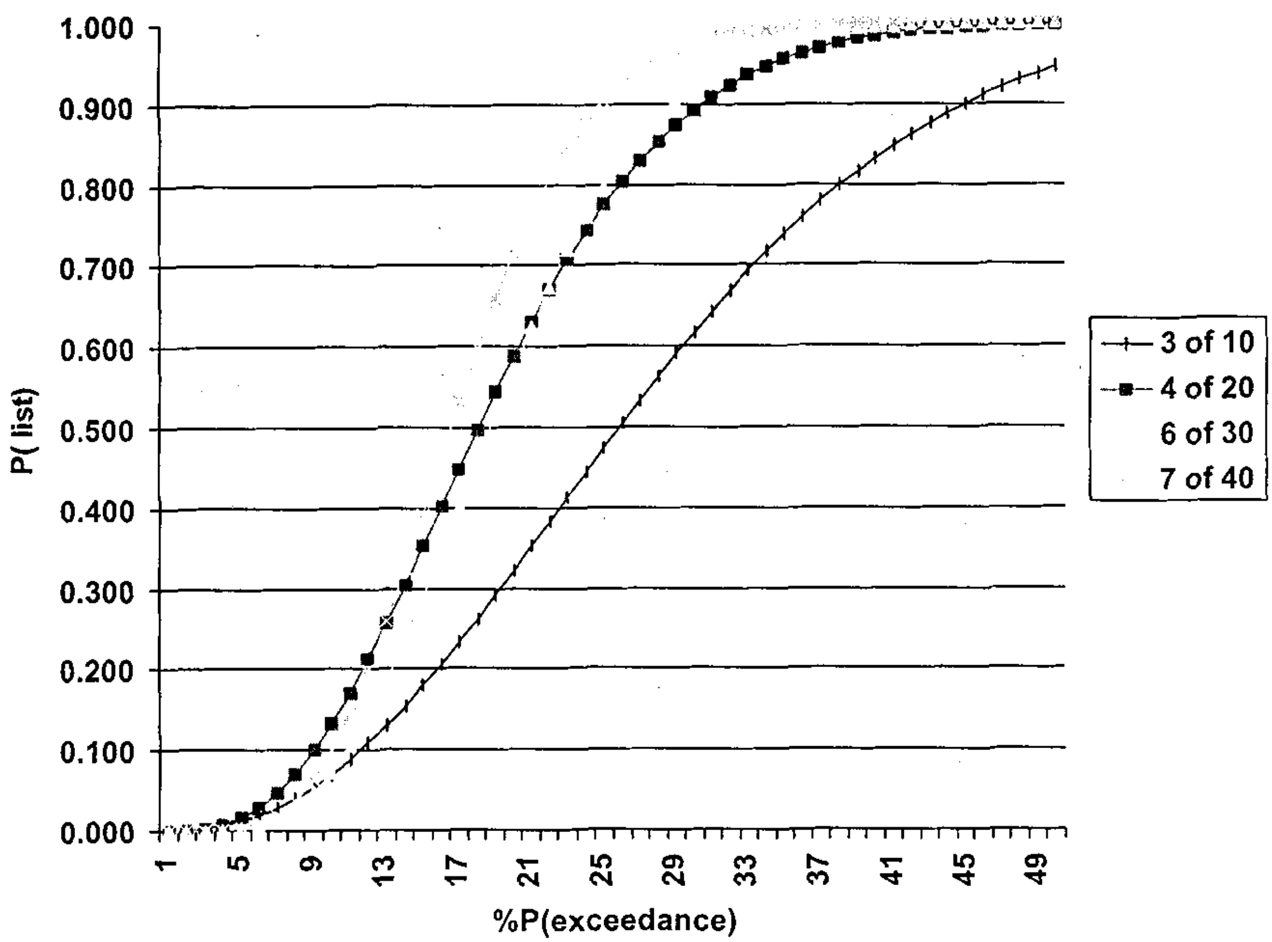
$$H_a\colon\ p > p_0 \tag{3.9}$$

at the $(100\alpha)\%$ level of significance, where p_0 is a number between 0 and 1 to be determined by the regulatory agency. For example, CRITBINOM(10, 0.10, 0.9298) returns the number "two", which means that $P(X \le 2 \mid n = 10,\ p = 0.1) \ge 0.9298$, i.e., the chance that the number of exceedances is two or less, given the exceedance probability of $p = 0.10$ and a sample size of $n = 10$, is at least 92.98%, and two is the smallest number of exceedances with this property. Therefore, when $p = 0.10$ (or less) the chance of three or more exceedances is less than 7.02%. Other examples can be generated similarly. Some are given below:

CRITBINOM(10, 0.1, 0.95) = 3,
CRITBINOM(15, 0.1, 0.95) = 4,
CRITBINOM(20, 0.1, 0.95) = 4,
CRITBINOM(30, 0.1, 0.95) = 6, and
CRITBINOM(40, 0.1, 0.95) = 7.

While Table 2 provides, for each n, $1 \le n \le 100$, the smallest number of exceedances x required for listing, it is important to calculate the probability of listing, $P(X \ge x \mid n, p)$, for each n and for various values of the true exceedance probability p. Table 3 gives the probabilities of listing for four sample sizes: $n = 10$, 20, 30, and 40 with p ranging from 0.01 to 0.50. These probabilities are plotted against the true exceedance probabilities in Chart 1, where the x-axis represents the true exceedance percentages $(100p)\%$ and the y-axis represents the probabilities of listing. Based on Chart 1, if the true exceedance probability of a pollutant at a particular water body is 0.1 (or less) and the proposed listing procedure is used, the chance of this water body reach being listed as impaired is (1) no more than 7% if only ten samples are collected, (2) no more than 13.3% if only 20 samples are collected, (3) no more than 7.3% if 30 samples are collected, and (4) no more than 9.95% if 40 samples are collected. If, on the other hand, the true exceedance probability of a pollutant at a water body is 0.25, then the chances of listing the water body as impaired with 10, 20, 30, and 40 samples are 47.4%, 77.5%, 79.7%, and 90.4%, respectively. It should be noted that, in the context of testing the null hypothesis H_0: $p \leq 0.1$ versus the alternative H_a: $p > 0.1$, the probability plots are actually the power curves for the four sample sizes. For each curve, i.e., for each sample size, the power of the test is an increasing function of the true exceedance probability, p. However, the four curves cross one another at some values of p. Thus, it is not necessary true that the larger the sample size is, the higher the probability of listing will be. For example, when the true exceedance probability is 0.1, the probability of listing is smaller for 30 samples with 6 exceedances than for 20 samples with 4 exceedances. The exact probabilities for both 20 and 30 samples when $p = 0.1$ can be found in Table 3.

Chart 1. Exceedance: 10% to List
With 95% Confidence Level



4. Delisting Procedure.

The problem of deciding by a statistical procedure whether or not to delist a body of water that has already been designated as "impaired" is not the same thing as deciding to list an impaired water. If the water body reach is no longer impaired, the regulator would want to be sure to delist it. On the other hand, if the water body reach is still impaired, the regulator would want to be sure to avoid delisting it. However, using a statistical procedure, no decision based on n sample measurements can be free from error; there will always be some chance of making a wrong decision. A sound statistical procedure is one that will minimize the chance of making a wrong decision.

In this section, it is assumed that a water body reach has been listed as impaired due to exceedances of a water quality criterion for a particular pollutant such as fluoride. Suppose that "$p < p_0$" is chosen as the method for delisting a water body reach due to an exceedance of a water quality criterion, where p_0 is a number between 0 and 1 to be determined by the regulatory agency. That is, an impaired water body, listed due to an exceedance, will be delisted whenever the true exceedance probability of the pollutant is less than p_0. The regulatory agency may consider using (1) $p_0 = 0.1$ or (2) $p_0 = 0.15$ or any other candidate value for delisting. A statistical procedure for delisting an impaired water body reach, due to an exceedance of a water quality criterion should provide the maximum number of exceedances, x, of the pollutant out of n sample measurements, allowed for the statistical conclusion "$p < p_0$" to be made with a high level of confidence. This can be achieved by the use of a hypothesis testing approach. The procedure is equivalent to rejecting the null hypothesis

$$H_0\colon\ p \geq p_0 \quad \text{(i.e., the water body is impaired),} \tag{4.1}$$

and accepting the alternative hypothesis

$$H_a\colon\ p < p_0 \quad \text{(i.e., the water body is not impaired).} \tag{4.2}$$

(Note that the null and alternative hypotheses for delisting are completely opposite to those of the listing procedure given in (3.6) and (3.7) for $p_0 = 0.1$.) The most powerful test is to reject the null hypothesis, at the $100\alpha\%$ (e.g., 5%) significance level, whenever the number of exceedances is less than or equal to x, where x satisfies the probability inequality:

$$P(X \leq x \mid n, p = p_0) \leq \alpha. \tag{4.3}$$

The number x obtained from (4.3) is the maximum number of exceedances, out of n sample measurements, allowed for delisting a water body reach with $(1-\alpha)100\%$ confidence. In the following, both options (1) $p < 0.1$ and (2) $p < 0.15$ for delisting an impaired water body are considered.

(1) Assume that the regulatory agency decides to use "$p < 0.1$" (i.e., $p_0 = 0.1$) as the delisting method. Then, for example, when $n = 28$, $p_0 = 0.1$, and $\alpha = 0.05$, the maximum number of exceedances is found to be $x = 0$. Equivalently, the Microsoft Excel function CRITBINOM(28, 0.1, 0.05) = 0, yielding the same result. For different sample sizes, and $p_0 = 0.1$, the maximum

number of exceedances, x, which are allowed for the acceptance of the alternative hypothesis $H_a: p < 0.1$ with approximately 95% confidence, are calculated using the Excel function and the results are given in Table 4. Based on the above calculation, when there is no exceedance among $n = 28$ measurements for a pollutant, we are 94.8% (or **approximately** 95%) confident that the true exceedance probability of the pollutant is below 0.1 and the water body will be removed from the impaired water list. Here, $n = 28$ is the smallest sample size that enables us to assess whether or not the true exceedance probability is below 0.1 with approximately (and closest to) 95% confidence. It is noted that the same conclusion should be reached using a lower one-sided 95% confidence interval approach. However, when inequality (3.3) is applied with $n = 28$, $x = 0$, and $\alpha = 0.05$, the smallest p is found to be U = 0.1045 giving the lower one-sided 95% confidence interval as [0, 0.1045], i.e., $p \leq 0.1045$. Notice that a minor discrepancy exists between the two results using the same data. This is because, under the hypothesis testing approach, "$p < 0.1$" is used for delisting with 94.8% confidence, and, under the confidence interval approach, "$p < 0.1045$" is used for delisting with 95% confidence. The exact 95% level of confidence cannot be accomplished if "$p < 0.1$" is to be used for delisting an impaired water. This is due to the fact that we are rounding off to the nearest whole number of exceedances. But for all practical purposes, both approaches provide the same conclusion with **approximately 95% confidence**.

For any sample size n less than or equal to 27, the level of confidence will be less than 95%. For example, when there is no exceedance among n = 10, 15, 20, and 25 sample measurements, the confidence levels are 65.13%, 79.41%, 87.84%, and 92.82%, respectively. Thus, $n = 28$ is the smallest sample size that is recommended for delisting with approximately 95% confidence.

Chart 2 plots the probabilities of delisting water body reaches with different true exceedance probabilities when 28 and 45 samples are collected. When the true exceedance probability of a pollutant at a particular water body is 0.01 (or less), the chances of delisting the water body reach based on 28 and 45 samples are 75.5% and 63.6%, respectively. When the true exceedance probability is 0.15, the delisting probabilities using the two sample sizes are 0.011 and 0.001, respectively. The delisting probabilities for 28 and 45 sample sizes for water body reaches with true exceedance probabilities between 0.01 and 0.25 are given in columns 2 and 3 of Table 5.

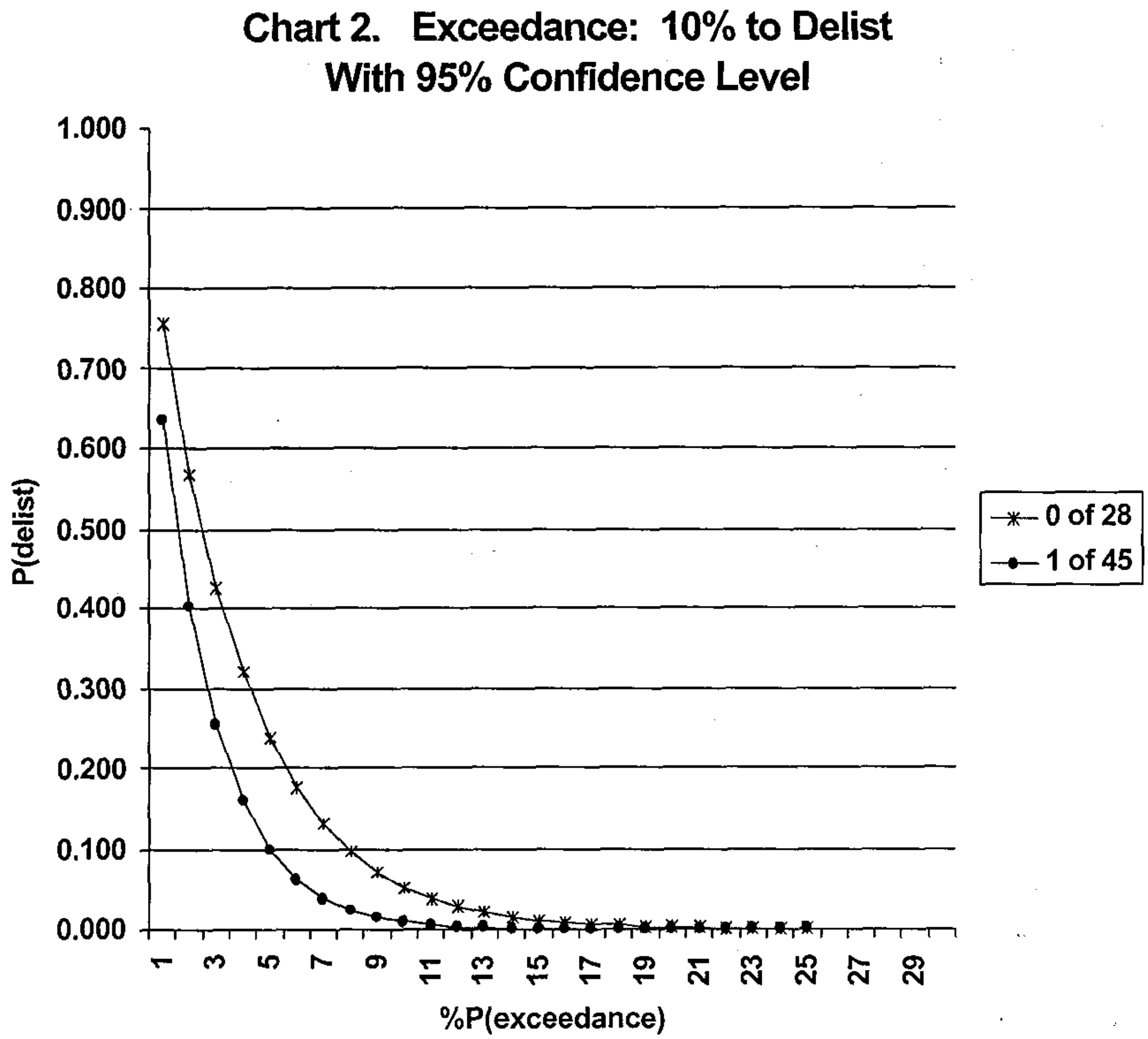


Chart 2. Exceedance: 10% to Delist
With 95% Confidence Level
1.000
0.900
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
P(delist)
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
%P(exceedance)
0 of 28
1 of 45

(2) Now, assume that the regulatory agency decides to use "$p < 0.15$" (i.e., the less-than-15%) method for delisting. Based on the calculation using (3.3), when there is no exceedance among 18 measurements, we can claim that, with 95% confidence, the true exceedance probability is below 0.15. Here, $n = 18$ is the smallest sample size that enables us to assess whether the true exceedance probability is below 0.15, with approximately (and closest to) 95% confidence. Similarly, when the sample size $n = 29$ and with only one exceedance in the 29 measurements, we are approximately 95% confident that the true exceedance probability is below 0.15 and the water body will be removed from the impaired water list. For different sample sizes, the maximum numbers of exceedances, x, for which we are approximately 95% confident that the true exceedance probability is less than 0.15, are also given in Table 4.

Assuming the less-than-15% method for delisting, Chart 3 plots the probabilities of delisting waterbody reaches with different true exceedance probabilities when 18 and 29 samples are collected. When the true exceedance probability of a pollutant at a particular water body is 0.01, the chances of delisting the water body reach for the 18 and 29 samples are 83.5% and 96.6%, respectively. When the true exceedance probability is 0.2, the delisting chances drop significantly to 1.8% and 1.3%, respectively. The delisting probabilities for 18 and 29 samples with true exceedance probabilities between 0.01 and 0.25 are given in columns 5 and 6 of Table 5, respectively.

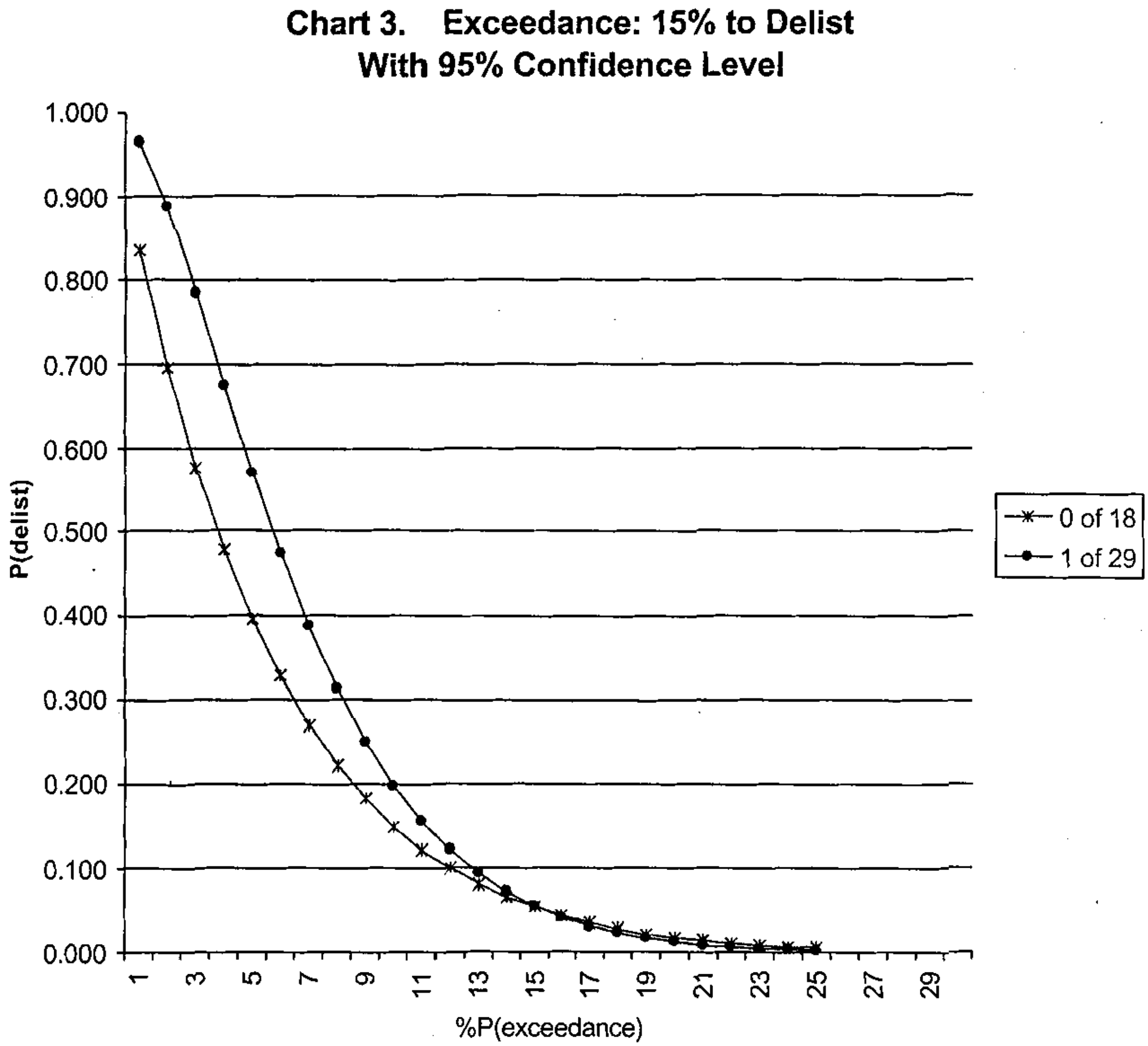


Chart 3. Exceedance: 15% to Delist
With 95% Confidence Level
P(delist)
1.000
0.900
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
%P(exceedance)
0 of 18
1 of 29

4. Conclusions and Discussion

In this study, we propose a nonparametric procedure for identifying impaired water body reaches in Florida based on the binomial distribution theory. The confidence interval approach and hypothesis testing approach are recommended for assessing the exceedance probability of a particular pollutant over its criterion. The starting premise for the listing procedure is that the water body should be listed if its true exceedance probability p of a pollutant is over 0.1. The decision to list an impaired water will be based on the minimum number of exceedances, x, found in n sample measurements. The minimum numbers required for listing are given in Table 2. For the delisting procedure, we provide two options depending on the true exceedance probability p: (1) $p < 0.1$ or (2) $p < 0.15$. Table 4 provides the maximum numbers of exceedances allowed for the water body reach to be removed from the impaired water list with approximately 95% confidence for both $p < 0.1$ and $p < 0.15$ options. In addition to the listing and delisting methods given in Tables 2 and 4, a table of listing probabilities and a table of delisting probabilities are provided. Also, three charts are presented showing the listing and delisting probabilities for selected sample sizes with different true exceedance probabilities.

In concluding this study, the issues on sample size, and on spatial and temporal coverage of samples are addressed below.

Sample Size. Because of limited sources and limited resources, the currently available samples for the majority of Florida water body reaches are quite small. (See, e.g., Table 1.) When estimating the true exceedance probability of a pollutant or testing hypotheses about the true exceedance probability, small sample sizes are associated with large uncertainty. For the proposed listing procedure, we suggest that ten or more sample measurements (minimum sample size $n = 10$) be required for assessing whether or not a water body reach is impaired based on criterion exceedances. The proposed delisting procedure requires stronger evidence and more information from sample than the listing procedure, if the same level of confidence is required. In order to assess whether or not the exceedance percentage of a pollutant in a particular water body is less than 10% for delisting, with approximately 95% confidence, we recommend that 28 or more water samples be collected for analysis.

The numbers of water samples required for the proposed listing and delisting procedures are different. Requiring "more samples" for delisting than for listing an impaired water at a comparable level of confidence seems somewhat puzzling to many readers, but it is strictly a matter of statistical theory. For example, suppose the agency decides that if p is shown to be greater than 0.1 then the water body will be listed as impaired. Assuming a null hypothesis of $p = 0.1$, the variance of each observation is $0.1 \times 0.9 = 0.09$. Now suppose the water body is listed as the result of a random sample. Then the agency will assume a null hypothesis of 0.2 for the purpose of testing for delisting. Now the variance of each observation is $0.2 \times 0.8 = 0.16$. Since the sample size necessary *to create the same level of confidence* for the estimation of p is roughly inversely proportional to the variance of an observation in the random sample, it will take more observations to provide the same standard of proof when $p = 0.2$ as when $p = 0.1$.

Consequently, it is not possible to use the same sample size to list and delist an impaired water body reach at the same level of confidence using the 10%-exceedance method for both listing and delisting. However, the same sample size could be used for listing and delisting at the expense of a lesser confidence level for delisting. As already demonstrated, we may use $n = 10$ samples for both listing and delisting. With three exceedances, the water body reach is listed as impaired with 92.98% confidence (from Table 2), while with no exceedance observed, out of the ten sample measurements, the water body is removed from the impaired water list with only 65.13% confidence (from Table 4). However, any statistical conclusion that has a confidence level of less than 90% is considered not acceptable by most statistics practitioners.

Spatial and Temporal Coverage of Samples. It is well-known that the concentration levels of many pollutants and metals depend on spatial location and season, and some physical or chemical properties, such as dissolved oxygen, vary dramatically at different time periods during a day. Based on these observations, we recommend that the sample measurements of a water body reach be collected randomly and at reasonably spread locations across the water surface. They are to be collected with sufficient temporal separation to ensure independence. In this way, the samples will be independent and unbiased. The true water quality of the whole reach will likely be represented by the sample measurements.

In this study, various statistical scenarios for listing and delisting an impaired water body are presented. These results should provide sufficient information and strong probabilistic evidence for the regulatory agency to render their decisions on the setting of clear guidelines for listing and delisting.

5. Acknowledgment. This research was supported in part by Contract LAB015 with the Bureau of Laboratories, the Florida Department of Environmental Protection. The authors wish to thank the assistance received from Mrs. Lori Wolfe and Mr. Daryll Joyner during the preparation of this report.

References.

Bickel, P.J. and Doksum, K.A. (2001). *Mathematical Statistics: Basic Ideas and Selected Topics*. Vol. I. (Second Edition). Prentice Hall.

Blyth, C.R. and Still, H.A. (1983). Binomial confidence intervals. *Journal of American Statistical Association* **V. 78**, 108-116.

Gilbert, O.R. (1987). *Statistical Methods for Environmental Pollution Monitoring*. John Wiley & Sons, Inc.

Table 1: Sample sizes for six pollutants in Florida

Organic Nitrogen			Dissolved Oxygen		
No. of Samples	No. of Reaches	Percent	No. of Samples	No. of Reaches	Percent
1-10	400	47%	1-10	380	39%
11-20	190	22%	11-20	188	19%
21-30	76	9%	21-30	77	8%
31-40	75	9%	31-40	103	10%
41-50	27	3%	41-50	42	4%
51-60	14	2%	51-60	25	3%
61-80	15	2%	61-80	26	3%
71-80	9	1%	71-80	21	2%
81-90	12	1%	81-90	11	1%
91-100	2	0%	91-100	11	1%
>100	29	3%	>100	99	10%
Grand Total	849	100%	Grand Total	983	100%

Total Nitrogen			Total Phosphorus		
No. of Samples	No. of Reaches	Percent	No. of Samples	No. of Reaches	Percent
1-10	373	36%	1-10	371	36%
11-20	194	19%	11-20	189	18%
21-30	98	9%	21-30	100	10%
31-40	110	11%	31-40	108	10%
41-50	37	4%	41-50	36	3%
51-60	24	2%	51-60	24	2%
61-80	40	4%	61-80	42	4%
71-80	28	3%	71-80	26	3%
81-90	20	2%	81-90	19	2%
91-100	18	2%	91-100	18	2%
>100	99	10%	>100	104	10%
Grand Total	1041	100%	Grand Total	1037	100%

NH4 (Ammonia)			NO3 (Nitrate)		
No. of Samples	No. of Reaches	Percent	No. of Samples	No. of Reaches	Percent
1-10	413	49%	1-10	388	37%
11-20	197	23%	11-20	197	19%
21-30	78	9%	21-30	72	7%
31-40	58	7%	31-40	82	8%
41-50	27	3%	41-50	24	2%
51-60	10	1%	51-60	16	2%
61-80	15	2%	61-80	16	2%
71-80	9	1%	71-80	10	1%
81-90	7	1%	81-90	10	1%
91-100	4	0%	91-100	4	0%
>100	24	3%	>100	26	3%
Grand Total	842	99%	Grand Total	845	81%

Table 2: To list a waterbody as impaired

With about 95% confidence, the minimum number of exceedances where you are sure the percentage of exceedances is greater than 10%

Sample Size n	*# exceedances*	*Conf Level %*	*Sample Size n*	*# exceedances*	*Conf Level %*
1	1	89.99	51	9	93.54
2	1	80.98	52	9	92.85
3	1	72.88	53	9	92.12
4	2	94.76	54	9	91.34
5	2	91.84	55	9	90.52
6	2	88.55	56	9	89.65
7	2	85.01	57	10	94.49
8	2	81.82	58	10	93.92
9	3	94.57	59	10	93.31
10	3	92.98	60	10	92.65
11	3	91.02	61	10	91.97
12	3	88.89	62	10	91.24
13	3	86.58	63	10	90.47
14	3	84.13	64	11	94.81
15	4	94.43	65	11	94.3
16	4	93.14	66	11	93.75
17	4	91.71	67	11	93.17
18	4	90.15	68	11	92.56
19	4	88.47	69	11	91.91
20	4	86.67	70	11	91.29
21	5	94.77	71	11	90.51
22	5	93.76	72	12	94.67
23	5	92.66	73	12	94.18
24	5	91.46	74	12	93.66
25	5	90.17	75	12	93.11
26	5	88.78	76	12	92.53
27	5	87.3	77	12	91.91
28	6	94.48	78	12	91.27
29	6	93.6	79	12	90.6
30	6	92.65	80	13	94.58
31	6	91.63	81	13	94.11
32	6	90.52	82	13	93.62
33	6	89.35	83	13	93.1
34	6	88.1	84	13	92.55
35	7	94.46	85	13	91.97
36	7	93.69	86	13	91.37
37	7	92.86	87	14	90.74
38	7	91.97	88	14	90.08
39	7	91.02	89	14	94.09
40	7	90.01	90	14	93.62
41	7	88.94	91	14	93.13
42	8	94.58	92	14	92.61
43	8	93.9	93	14	92.06
44	8	93.18	94	14	91.49
45	8	92.4	95	15	90.9
46	8	91.56	96	15	90.28
47	8	90.68	97	15	94.1
48	8	89.75	98	15	93.66
49	9	94.79	99	15	93.19
50	9	94.18	100	15	92.69

Table 3. Listing Probabilities

(Using Greater-than-10% Exceedance for Listing)

Exceedance	Listing Probabilities			
Prob.	3 of 10	4 of 20	6 of 30	7 of 40
0.01	0.000	0.000	0.000	0.000
0.02	0.001	0.001	0.000	0.000
0.03	0.003	0.003	0.000	0.000
0.04	0.006	0.007	0.001	0.001
0.05	0.012	0.016	0.003	0.003
0.06	0.019	0.029	0.008	0.009
0.07	0.028	0.047	0.016	0.020
0.08	0.040	0.071	0.029	0.038
0.09	0.054	0.099	0.048	0.064
0.1	0.070	0.133	0.073	0.100
0.11	0.088	0.171	0.105	0.144
0.12	0.109	0.213	0.143	0.198
0.13	0.131	0.257	0.187	0.259
0.14	0.155	0.304	0.236	0.324
0.15	0.180	0.352	0.289	0.393
0.16	0.206	0.401	0.345	0.463
0.17	0.234	0.450	0.403	0.532
0.18	0.263	0.497	0.461	0.597
0.19	0.292	0.544	0.517	0.658
0.2	0.322	0.589	0.572	0.714
0.21	0.353	0.631	0.625	0.764
0.22	0.383	0.671	0.674	0.808
0.23	0.414	0.708	0.719	0.845
0.24	0.444	0.743	0.760	0.877
0.25	0.474	0.775	0.797	0.904
0.26	0.504	0.804	0.830	0.925
0.27	0.534	0.830	0.859	0.943
0.28	0.562	0.853	0.884	0.957
0.29	0.590	0.874	0.905	0.968
0.3	0.617	0.893	0.923	0.976
0.31	0.643	0.909	0.939	0.983
0.32	0.669	0.923	0.951	0.988
0.33	0.693	0.936	0.962	0.991
0.34	0.716	0.946	0.970	0.994
0.35	0.738	0.956	0.977	0.996
0.36	0.759	0.963	0.982	0.997
0.37	0.779	0.970	0.986	0.998
0.38	0.798	0.976	0.990	0.999
0.39	0.816	0.980	0.992	0.999
0.4	0.833	0.984	0.994	0.999
0.41	0.848	0.987	0.996	1.000
0.42	0.863	0.990	0.997	1.000
0.43	0.876	0.992	0.998	1.000
0.44	0.889	0.994	0.998	1.000
0.45	0.900	0.995	0.999	1.000
0.46	0.911	0.996	0.999	1.000
0.47	0.921	0.997	0.999	1.000
0.48	0.930	0.998	1.000	1.000
0.49	0.938	0.998	1.000	1.000
0.5	0.945	0.999	1.000	1.000

Table 4: To delist a waterbody from impaired

With 95% confidence, the maximum number of exceedances, x, where you are sure the percentage of exceedances is less than 10%

n	x	% Conf	n	x	% Conf
1			51	2	89.61
2			52	2	90.44
3			53	2	91.02
4			54	2	91.66
5			55	2	92.26
6			56	2	92.81
7			57	2	93.34
8			58	2	93.82
9			59	2	94.27
10	0	65.13	60	2	94.7
11	0	68.62	61	3	87.1
12	0	71.76	62	3	87.9
13	0	74.58	63	3	88.66
14	0	77.12	64	3	89.47
15	0	79.41	65	3	90.04
16	0	81.47	66	3	90.68
17	0	83.32	67	3	91.28
18	0	84.99	68	3	91.84
19	0	86.49	69	3	92.38
20	0	87.84	70	3	92.88
21	0	89.06	71	3	93.35
22	0	90.15	72	3	93.79
23	0	91.14	73	3	94.2
24	0	92.02	74	3	94.59
25	0	92.82	75	3	94.96
26	0	93.54	76	4	88.79
27	0	94.19	77	4	89.44
28	0	94.77	78	4	90.06
29	1	80.11	79	4	90.65
30	1	81.63	80	4	91.1
31	1	83.06	81	4	91.73
32	1	84.36	82	4	92.23
33	1	85.58	83	4	92.7
34	1	86.71	84	4	93.15
35	1	87.76	85	4	93.57
36	1	88.74	86	4	93.97
37	1	89.64	87	4	94.34
38	1	90.47	88	4	94.7
39	1	91.24	89	5	89.08
40	1	91.95	90	5	89.68
41	1	92.61	91	5	90.24
42	1	93.22	92	5	90.78
43	1	93.77	93	5	91.3
44	1	94.29	94	5	91.79
45	1	94.76	95	5	92.25
46	2	85.16	96	5	92.69
47	2	86.17	97	5	93.11
48	2	87.11	98	5	93.51
49	2	88	99	5	93.88
50	2	88.83	100	5	94.24

With 95% confidence, the maximum number of exceedances, x, where you are sure the percentage of exceedances is less than 15%

n	x	% Conf	n	x	% Conf
1			51	4	89.78
2			52	4	90.69
3			53	4	91.54
4			54	4	92.31
5			55	4	93.02
6			56	4	93.67
7			57	4	94.27
8			58	4	94.81
9			59	5	89.44
10	0	80.31	60	5	90.32
11	0	83.27	61	5	91.14
12	0	85.78	62	5	91.89
13	0	87.91	63	5	92.59
14	0	89.72	64	5	93.24
15	0	91.26	65	5	93.83
16	0	92.57	66	5	94.38
17	0	93.69	67	5	94.88
18	0	94.64	68	6	90.1
19	1	80.15	69	6	90.89
20	1	82.44	70	6	91.62
21	1	84.5	71	6	92.3
22	1	86.33	72	6	92.93
23	1	87.96	73	6	93.52
24	1	89.41	74	6	94.06
25	1	90.69	75	6	94.56
26	1	91.83	76	6	95.03
27	1	92.84	77	7	90.75
28	1	93.73	78	7	91.45
29	1	94.51	79	7	92.11
30	2	84.86	80	7	92.73
31	2	86.41	81	7	93.3
32	2	87.82	82	7	93.83
33	2	89.1	83	7	94.33
34	2	90.25	84	7	94.79
35	2	91.3	85	8	90.68
36	2	92.24	86	8	91.36
37	2	94.08	87	8	92
38	2	94.85	88	8	92.6
39	2	94.53	89	8	93.16
40	3	86.98	90	8	93.68
41	3	88.21	91	8	94.16
42	3	89.33	92	8	94.62
43	3	90.36	93	9	90.68
44	3	91.29	94	9	91.33
45	3	92.15	95	9	91.95
46	3	92.93	96	9	92.52
47	3	93.64	97	9	93.07
48	3	94.28	98	9	93.57
49	3	94.87	99	9	94.05
50	4	88.79	100	9	94.49

Table 5. Delisting Probabilities

	Less than 10% to delist	
Exceedance Prob.	0 of 28 (Delisting Prob.)	1 of 45 (Delisting Prob.)
0.01	0.755	0.925
0.02	0.568	0.773
0.03	0.426	0.607
0.04	0.319	0.458
0.05	0.238	0.335
0.06	0.177	0.239
0.07	0.131	0.167
0.08	0.097	0.115
0.09	0.071	0.078
0.1	0.052	0.052
0.11	0.038	0.035
0.12	0.028	0.023
0.13	0.020	0.015
0.14	0.015	0.009
0.15	0.011	0.006
0.16	0.008	0.004
0.17	0.005	0.002
0.18	0.004	0.001
0.19	0.003	0.001
0.2	0.002	0.001
0.21	0.001	0.000
0.22	0.001	0.000
0.23	0.001	0.000
0.24	0.000	0.000
0.25	0.000	0.000

	Less than 15% to delist	
Exceedance Prob.	0 of 18 (Delisting Prob.)	1 of 29 (Delisting Prob.)
0.01	0.835	0.966
0.02	0.695	0.886
0.03	0.578	0.784
0.04	0.480	0.676
0.05	0.397	0.571
0.06	0.328	0.474
0.07	0.271	0.388
0.08	0.223	0.314
0.09	0.183	0.251
0.1	0.150	0.199
0.11	0.123	0.156
0.12	0.100	0.122
0.13	0.082	0.094
0.14	0.066	0.072
0.15	0.054	0.055
0.16	0.043	0.042
0.17	0.035	0.031
0.18	0.028	0.023
0.19	0.023	0.017
0.2	0.018	0.013
0.21	0.014	0.009
0.22	0.011	0.007
0.23	0.009	0.005
0.24	0.007	0.004
0.25	0.006	0.003