2022 Integrated Report Assessment Methodology

Background paper: Minimum data requirements for category 2

Sept. 22, 2020

State of Oregon
Department of Environmental Quality

Water Quality
700 NE Multnomah St.
Suite 600
Portland, OR 97232
Phone: 503-229-5696
800-452-4011
Fax: 503-229-6124
Contact: Becky Anthony
integratedreport@deq.state.or.us

www.oregon.gov/DEQ

DEQ undertook significant improvements to its assessment methodologies in the 2018/2020 Integrated Report cycle. One of these improvements included implementation of the binomial hypothesis test in its listing and delisting methodologies for numeric water quality standards for both toxic substances and conventional pollutants. The binomial hypothesis test accounts for sample size, errors in sample accuracy and precision, and explicitly defines the acceptable levels of certainty in making a categorical determination.

Evaluating waterbodies with samples collected implies that the characteristics of the waterbody are accurately represented by the samples. Grab sampling inherently introduces bias, error, variability and uncertainty about how well the samples represent the waterbody as a whole over a given time. Statistical analysis of water quality samples provides a quantifiable way to describe the confidence that a waterbody attains or exceeds a water quality criteria (which is comprised of magnitude, duration and frequency metrics) based on the samples collected.

Use of the binomial hypothesis test for the assessment of water quality defines the risk of making errors in determining both impairment and attainment, and the risks can be weighed for supporting a given conclusion. DEQ's focus in the 2018/2020 methodology update was to strengthen its method for determining impairment of a waterbody by using a statistical method. Conversely, DEQ implemented the binomial method to remove waterbodies from the impaired waters list and increase confidence that water quality samples reflected the current state of a restored waterbody. However, this update set a default error rate of 64% for incorrectly identifying a waterbody as attaining water quality standards. DEQ determined this is not an acceptable error rate for classifying waters as supporting their beneficial uses.

Problem: Category 2 identifies waterbodies where beneficial uses are supported and water quality criteria are being attained. During the 2018/2020 assessment process, a minimum sample threshold based on the binomial statistical method was never defined for attainment purposes. Currently, as little as two samples (toxic pollutants) and five samples (conventional pollutants) in a Period of

Record (POR = 5 to 10 years) can constitute attainment and are considered "fully supporting" their beneficial use which increases the probability of Type II errors (failing to identify a true impairment). This methodology update would clarify minimum sample requirements for what constitutes "attainment" of a criteria based on the binomial hypothesis test, a statistically-based methodology.

Decision error, or an incorrect conclusion of waterbody status can occur when the collected sample data, may be unrepresentative of the population as a whole. For example, in a waterbody assessment, the assumption of the null hypothesis to be tested is: The water body is meeting water quality standards. If this hypothesis is correct (i.e., the water body is meeting water quality standards) and the statistical analysis leads to that conclusion, then a correct decision to not reject the null hypothesis is made. Therefore, beneficial uses are supported and the water body will not be recommended for placement on the section 303(d) list of impaired waters.

Conversely, the samples, by chance, can indicate a greater degree of impairment in the particular samples than actually occurs across the waterbody as a whole. In that case, the samples would not represent the true population and, an erroneous conclusion would be made that the assessment unit as a whole does not meet water quality standards. Using the assessment methodology, the null hypothesis would be rejected and the water would mistakenly be recommended for placement on the section 303(d) list. This is an example of a Type I error, incorrectly rejecting a true null hypothesis (Table 1).

However, if the null hypothesis is false (i.e., the water is impaired) an error can also be made if the sample data, suggests that the water body is not impaired when, in fact, it is. This is called a Type II error (failing to reject an untrue null hypothesis). Type II errors can occur when appropriate minimum sample sizes are not defined.

Decision	**Reality**	
	H_0 is true i.e. Waterbody is not impaired	**H_0 is false i.e. Waterbody is impaired**
Reject H_0	Type I (false positive) error	Correct decision
Do not reject H_0	Correct decision	Type II (false negative) error

Table 1. The two types of statistical error

Most basic statistical tests only allow direct control (i.e., limitation) over Type I error rates through the formulation of the null hypothesis. Type II error rates, however, may be lowered or controlled most effectively by increasing sample size, increasing effect size, or decreasing the overall range/distribution of sample values.

Since the overall range or distribution of sample values cannot be controlled as part of the assessment and the effect size of 15% was established in the 2018/2020 methodology, DEQ has retained the current assessment methodology and selected to control Type II error rates by increasing the minimum sample size to make a Category 2 determination.

The power (1 – β) in the binomial statistical method lies in the ability to differentiate between a water body that supports its beneficial uses (i.e. Attaining) versus a waterbody that does not (i.e. Impaired). By maximizing the power in the statistical test, DEQ has chosen to minimize Type II errors of identifying a waterbody as supporting its uses when it is in fact, impaired.

n	k_i	α	β	|α - β|	C.I. (1-α)	Power (1-β)
2	1	0.0975	0.64	0.5425	0.9025	0.36
3	1	0.14263	0.512	0.36938	0.85738	0.488
4	1	0.18549	0.4096	0.22411	0.81451	0.5904
5	1	0.22622	0.32768	0.10146	0.77378	0.67232
6	1	0.26491	0.26214	0.00276	0.73509	0.737856
7	1	0.30166	0.20972	0.09195	0.69834	0.7902848
8	1	0.33658	0.16777	0.16881	0.66342	0.8322278
9	1	0.36975	0.13422	0.23553	0.63025	0.8657823
10	1	0.40126	0.10737	0.29389	0.59874	0.8926258

Figure 1. Binomial calculator for Aquatic Life toxics

n	k_i	α	β	|α - β|	C.I. (1-α)	Power (1-β)
2	1	0.19	0.5625	0.3725	0.81	0.4375
3	1	0.271	0.421875	0.150875	0.729	0.578125
4	1	0.3439	0.316406	0.027494	0.6561	0.68359375
5	1	0.40951	0.237305	0.172205	0.59049	0.762695313
6	1	0.468559	0.177979	0.29058	0.531441	0.822021484
7	1	0.521703	0.133484	0.388219	0.478297	0.866516113
8	1	0.569533	0.100113	0.46942	0.430467	0.899887085

Figure 2. Binomial calculator for conventional pollutants

Recommendation:

	Minimum sample size for Category 2	Type II error rate (β)	Power (1 – β)
Aquatic Life Toxics	10	10%	~ 90%
Conventional Pollutants	8	10%	~ 90%

Table 2. Minimum sample size recommendations for Category 2

DEQ selected a Type II error rate (β) of approximately 10% which corresponds to 90% power value (1 – β) to support its minimal sample size recommendations for both Aquatic Life toxics criteria and conventional pollutants (Figures 1 and 2; Table 2). This is consistent with the Type I error rate chosen for impairment and implemented in the 2018 Water Quality Assessment Methodology.