T3

rec'd
1/19/04
Formal
Peer
Reviewer.

From: John Rice <rice@stat.berkeley.edu>
To: <lacaf@dwq.swrcb.ca.gov>
Date: 1/20/04 12:13PM
Subject: review of draft policy

Dear Mr. LaCaro,

Please find attached my review of the draft of Water Quality Control Policy for Developing California's Clean Water Act Section 303(d) List.

I would be glad to discuss this review with you or other interested parties.

Sincerely yours,

John Rice

....

John Rice
Department of Statistics
University of California, Berkeley 367 Evans Hall, Berkeley, CA. 94720-3[illegible]
http://www.stat.berkeley.edu/users/rice

CC: <flocdoc@pacbell.net>, <scott.macdonald@ucop.edu>

Review:
Functional Equivalent Document: Water Quality Control Policy for Developing California's Clean Water Act Section 303(d) List
December 2003 Draft

John A. Rice
Professor of Statistics
University of California, Berkeley
January 20, 2004

My expertise and my comments on this document relate to Issue 6: Statistical Evaluation of Numeric Water Quality Data.

Issue 6A: Selection of Hypotheses to Test

The discussion in this section conforms with standard statistical theory, correctly stating that standard statistical tests can only explicitly control the probability of Type I error and tracing the consequences for two alternatives.

Issue 6B: Choice of Statistical Tests

The pros and cons of several possible tests are discussed. The recommendation of the exact binomial test for reasons of simplicity and robustness seems reasonable. As is pointed out, the major shortcoming of this test is that it does not take magnitude into account. A key parameter in this test is the critical exceedance rate (CER). This parameter can be interpreted in the following way: the actual sample of values upon which the test is based are viewed as a simple random sample from a large, possibly infinite, population of such samples. If in this population the proportion of samples greater than the nominal standard is greater than CER, the hypothesis that the body of water achieves water quality standards is false; otherwise it is true.

Type I and II error rates are displayed in Figures 15 and 16, but some crucial information was missing in Figure 16. Namely, the chance of a Type II error depends upon the population value of CER, and the value which was used in constructing Figure 16 was not given (unless I somehow overlooked it in the text).

Issue 6C: Selection of Statistical Confidence Level

I found the terminology in this section a little confusing. The statement, "Statistical confidence is the probability that a hypothesis is true," is only valid under a Bayesian framework. In the "frequentist" framework in which hypothesis testing is done, controlling Type I error rates, this statement is not literally true—a hypothesis is either true or false and no probability can be attached to it.

However, the document does not use this interpretation of statistical confidence level, but takes this phrase to mean $1-\alpha$. The terminology may be a little nonstandard; it would have been more standard to title the section "Selection of Statistical Significance Level," since α is commonly referred to as the significance level of the test. However, this may be nit-picking.

Decreasing α will generally increase β, so some statements on page 164 could be stronger. As α is decreased it is harder to reject the null hypothesis.

Since the choice of α is, from a purely mathematical point of view, arbitrary, it must be selected by considering the subject matter, and the recommendation of this section seems defensible to me, although I am not sufficiently expert in the subject matter to have a strong opinion.

Issue 6D: Critical Rate of Exceedances of Water Quality Standards

This section discusses the choice of CER. Frankly, I found the writing somewhat confusing, because, throughout, it did not explicitly differentiate whether the "proportion of samples" refers to those samples actually conducted to assess water quality or to the hypothetical proportion in the population (see above). This potential confusion between actual sample data and population data pervades the discussion. Thus for example, the definition, "The *critical exceedance rate* is the proportion of samples that exceed an applicable water quality criterion ("the proportion of exceedances") providing overwhelming evidence that a water segment fails to meet water quality standards for a particular pollutant," is a little ambiguous. The second and third paragraphs of the Issue Description are similarly confusing; for example, it does not seem to me to be appropriate to include items 1 and 6 in the list of sources of uncertainty relevant to setting the CER. They have to do with the sample size and the test, not the population. Also the statement that "the critical rate of exceedance will determine the Type II error" is somewhat misleading, since the *definitions* of the null and alternative hypotheses depend upon the specification of CER.

These objections are, however, more semantic than substantive. The case for using CER greater than zero is clear, when one considers measurement error, sample unit definition, averaging period, etc.

I was confused by the paragraph on page 173 that begins, "To determine…" It seems to me that it is easier to reject a null hypothesis that specifies a small CER than it is to reject one that specifies a larger value, since the former test rejects whenever the latter one does. Thus, I don't understand the claim of the following paragraph either.

Issue 6E: Minimum Sample Size

As is clearly stated, the critical issue is that, with Type I error rate and CER held constant, the probability of a Type II error depends upon the sample size, decreasing as the sample size increases.

The discussion of Alternative 2 refers to Figure 15, but I think that this is a typo and that it should refer to Figure 16. Even so, the particular alternative under which the Type II error rate is graphed in Figure 16 needs to be specified. (The Type II error rate depends upon what the alternative exceedance rate is). In the last sentence of that paragraph, "sample population of 10" should presumably read "sample size of 10," and the particular alternative exceedance rate should be specified.

The discussion of Alternative 3 states, "Smaller sample sizes are more prone to yield erroneous decisions to list." This is not true if the null hypothesis is that the water segment meets water quality standards, because the chance of an erroneous decision to list is controlled by α, which is set in advance.

The paragraph on page 180 that begins, "While Type II error…" is confusing.

The last paragraph under Alternative 3 is not quite right in detail, although it is in spirit. By my calculation, if $\alpha=0.10$ and $n=22$, the decision to list would require five or more exceedances, while the decision to delist would require zero exceedances, when the exact binomial test is used. If the sample size were less than 22, it would be impossible to conduct an exact binomial test to delist with $\alpha=0.10$.

The last sentence of the first paragraph of the discussion of Alternative 4 does not make sense to me.

The gist of the issue is that small sample sizes provide little power, making it unlikely to detect violations of standards, unless the exceedance rates are very large. Similarly, the chance of legitimate delisting is small, unless the exceedance rates are very small. The probability of type I error, α, is always controlled by the hypothesis testing procedure. I don't find the single (and ambiguous) Figure 16 sufficient for deciding these important policies on sample size. More extensive calculations of power as a function of true exceedance rate need to be done for a range of plausible sample sizes, and for both listing and delisting. And these calculations are straightforward to do.

Issue 6F: Quantitation of Chemical Measurements

As the document states, one of the advantages of the exact binomial test is that there is no ambiguity in how to treat measurements below the quantitation limit, so long as that limit is less than the water quality objective. When the quantitation limit is larger than the water quality objective, measurements between the two are indeed difficult to interpret. The labeling of Figure 22 is incomplete (the upper horizontal line should be labeled QL and the lower WQO).

Received: from swrcb10.swrcb.ca.gov
by gwgate.swrcb.ca.gov; Tue, 20 Jan 2004 12:13:25 -0800
Received: from sccrmhc12.comcast.net (sccrmhc12.comcast.net [204.127.202.56])
by swrcb10.swrcb.ca.gov (8.12.10+Sun/8.12.10) with SMTP id i0KKBsNF006849
for <lacaf@dwq.swrcb.ca.gov>; Tue, 20 Jan 2004 12:11:54 -0800 (PST)
Received: from [192.168.1.103] (c-24-7-97-106.client.comcast.net[24.7.97.106])
by comcast.net (sccrmhc12) with ESMTP
id <20040120201308012004q9fae>; Tue, 20 Jan 2004 20:13:09 +0000
Date: Tue, 20 Jan 2004 12:13:11 -0800
From: John Rice <rice@stat.berkeley.edu>
To: lacaf@dwq.swrcb.ca.gov
cc: flocdoc@pacbell.net, scott.macdonald@ucop.edu
Subject: review of draft policy
Message-ID: <13065697.1074600791@[192.168.1.103]>
X-Mailer: Mulberry/3.0.3 (Win32)
MIME-Version: 1.0
Content-Type: multipart/mixed; boundary="==========13071466=========="

From: John Rice <rice@stat.berkeley.edu>
To: Tim Stevens <STEVT@dwq.swrcb.ca.gov>
Date: 1/27/04 9:17AM
Subject: Re: Review of Draft 303(d) Listing Policy FED for SWRCB

Yes, you're right -- I meant "To distinguish"

John

--On Tuesday, January 27, 2004 9:10 AM -0800 Tim Stevens
<STEVT@dwq.swrcb.ca.gov> wrote:

> Dr. Rice,
>
> I am reviewing your comments* on the statistical portion of the State
> Water Resources Control Board's (SWRCB) draft 303(d) Listing Policy
> Functional Equivalent Document (FED). So far, I have one question. On
> Page 2 of your comments you state:
>
> "I was confused by the paragraph on page 173 that begins,
> 'To determine...' "
>
> I have examined that section of the draft FED (see attached WORD file)
> and was unable to find a paragraph that begins with "To determine."
> However, there is one that begins, "To distinguish" on Page 173. Is
> that the one you are referring to?
>
> Thank you for taking the time to clarify this comment, as we want to be
> sure to address all of your comments and to make revisions in the
> correct areas.
>
> Tim Stevens
> 916/341-5911
> ____________
> * We greatly appreciate your comments and believe that revisions based
> on them will improve the validity and accuracy of the FED.
>
>
>
>
>
> Timothy P. Stevens
> Environmental Scientist
> Division of Water Quality
> State Water Resources Control Board
> Street Address: 1001 I Street, Sacramento, CA 95814
> Mailing Address: P.O. Box 244213, Sacramento, CA 94244-2130
> 916/341-5911 (work phone)
> 916/341-5550 (FAX)

....
John Rice
Department of Statistics
University of California, Berkeley