FISCAL YEARS 1999 & 2000 REPORT OF THE WATER QUALITY ACADEMIC ADVISORY COMMITTEE

SPECIAL REPORT

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
BLACKSBURG, VIRGINIA

2000

Support for this project and the printing of this publication was provided by a grant from the Virginia Department of Environmental Quality and the Virginia Water Resources Research Center. The contents of this publication do not necessarily reflect the views or policies of the Virginia Department of Environmental Quality or the Virginia Water Resources Research Center. The mention of commercial products, trade names, manufacturers, or services does not constitute an endorsement or recommendation.

Additional copies are available while the supply lasts and may be obtained from the
Virginia Water Resources Research Center
10 Sandy Hall
Blacksburg, VA 24061
(540)231-5624
FAX: (540)231-6673
e-mail: water@vt.edu
homepage address: **http://www.vwrrc.vt.edu**
Single copies are free to Virginia residents.

Leonard A. Shabman, Director

FISCAL YEARS 1999 & 2000 REPORT OF THE WATER QUALITY ACADEMIC ADVISORY COMMITTEE

PREPARED FOR THE VIRGINIA DEPARTMENT OF ENVIRONMENTAL QUALITY, OFFICE OF WATER QUALITY PROGRAMS

LEONARD A. SHABMAN (CHAIR)
VIRGINIA WATER RESOURCES RESEARCH CENTER
VIRGINIA TECH

THEO DILLAHA
BIOLOGICAL SYSTEMS ENGINEERING
VIRGINIA TECH

CARL HERSHNER
RESOURCE MANAGEMENT & POLICY
VIRGINIA INSTITUTE OF MARINE SCIENCE

HOWARD I. KATOR
BIOLOGICAL SCIENCES
VIRGINIA INSTITUTE OF MARINE SCIENCE

SAIED MOSTAGHIMI
BIOLOGICAL SYSTEMS ENGINEERING
VIRGINIA TECH

ERIC P. SMITH
DEPARTMENT OF STATISTICS
VIRGINIA TECH

LEONARD A. SMOCK
BIOLOGY DEPARTMENT
VIRGINIA COMMONWEALTH UNIVERSITY

TAMIM YOUNOS
VIRGINIA WATER RESOURCES RESEARCH CENTER
VIRGINIA TECH

SHAW L. YU
CIVIL ENGINEERING & APPLIED SCIENCE

CARL E. ZIPPER
CROP AND SOIL ENVIRONMENTAL SCIENCE
VIRGINIA TECH

Virginia Water Resources Research Center
Virginia Tech
10 Sandy Hall
Blacksburg, VA 24061

May 2000

Special Report SR18-2000

Table of Contents

		Page
1.	**Introduction**	**5**
2.	**Watershed Management: Looking to The 21st Century** *Leonard Shabman*	**6**
3.	**Assessing Violations of Water Quality Standards Under Section 303 (d) of the Clean Water Act** *Eric P. Smith, Keying Ye, Chris Hughes, Leonard Shabman*	**11**
4.	**Virginia's Water Quality Standards: Future Development Issues and Recommendations** *Carl Hershner*	**39**
5.	**Review of the Commonwealth's Biological Monitoring Program** *Leonard Smock*	**51**
6.	**Bio-monitoring – The New Water Quality Management Challenge** *Leonard Smock*	**55**
7.	**Addressing the Issue of Fecal Indicators in Virginia's Waters** *Howard Kator*	**59**
8.	**Sampling and Monitoring Design for 305(d) Program – Combining Probabilistic and Judgmental Sampling** *Eric P. Smith*	**61**
9.	**Modeling and the Alternatives for Water Quality Management** *Theo Dillaha and Shaw Yu*	**63**
	Appendix A: *Proposed* Fiscal year 2001 Work Plan of the Water Quality Academic Advisory Committee	**66**

1. Introduction

The Virginia General Assembly in the 1997 *Water Quality Monitoring, Information and Restoration Act (WQMIR)* directed the DEQ to develop the EPA-required 303(d) and 305(b) reports in consultation with experts from the states' universities. Also, WQMIR requires the DEQ to "develop and publish a procedure governing its process for defining and determining impaired waters and shall provide for public comment on the procedure" with the assumption that these 303(d) procedures will be developed after consultation with scientists from states' universities. To meet WQMIR academic consultation requirements, DEQ asked the Virginia Water Resources Research Center (VWRRC) to organize and coordinate a Water Quality Academic Advisory Committee (WQAAC) as an independent advisory body.

In 1997-1998, the WQAAC reviewed and evaluated the scientific merits of the DEQ's existing and evolving water quality assessment procedures for the 305(b) and 303(d) reports. The WQAAC's 1997-1998 findings submitted to DEQ are documented in VWRRC Special Report No. SR8-1998. Hard copies of the report are available from the VWRRC. Electronic copies can be located at Website www.vwrrc.vt.edu. Overall, in its 1998 report, the WQAAC made 17 findings and recommendations to the DEQ. As an overall summary, the WQAAC did not recommend that DEQ make immediate changes to its assessment guidelines. However, the WQAAC did recommend a number of future actions.

This report documents activities of WQAAC for the fiscal years 1999 and 2000. The specific topics addressed during this time period were based on recommendations made by the WQAAC to the DEQ in 1997-1998 report and additional consultation with the DEQ personnel. *The Virginia Water Resources Research Center (VWRRC) coordinated the activities during this time period. AAC members worked in teams, however individual authors prepared the papers in this report. While all of the WQAAC members may agree with the papers, the responsibility for the content and conclusions rests with the individual authors.*

In 1998- 2000, the reports and papers included here were prepared by members of the WQAAC and address the topics of DEQ concern that were part of the WQAAC work plan. Some of the papers constitute findings of the work effort and some of the contents are letter reports on progress made in addressing different issues.

Overall in the report, three substantive areas are identified as common theme between all tasks: issues relating to biological monitoring protocols; modeling techniques and other methods relating stressors to benthic degradation and fecal contamination; and water quality goals and setting water quality standards. These issues will constitute further WQAAC deliberations in 2000-2001.

2. Watershed Management: Looking to The 21st Century

Leonard Shabman
Director, Virginia Water Resources Research Center

Remarks prepared for presentation at the Virginia Department of Environmental Quality
Office of Water Quality Programs
Retreat at Marion, VA, May 18, 1999

Point 1: Have Perspective

Early in this century Cockrells Creek, near my summer cottage on the Chesapeake Bay, was a dead zone. The creek received the effluent of tomato packing plants and poultry and fish processing waste. Wooden pilings for docks lasted for decades because ship worms could not live in the creek's waters. Today the fish processing continues, but environmental laws make effluent treatment an unavoidable requirement. And today Cockrells Creek supports a viable blue crab fishery, extensive water contact recreation -- and pier pilings are now attacked by ship worms!

A few years ago the Water Center sponsored a undergraduate student's canoe trip along the length of the Roanoke River to monitor water quality. Among the interesting statements in his trip report was that in a historical context river water quality was in better condition today than the recent past.

These impressions about progress on water quality are consistent with Carl Zipper and Golde Holtzman's trends assessment – that you will hear about later on. After we recognize the Commonwealth's remarkable economic and population growth we must be pleased that Virginia water quality – as measured by some of the traditional chemical indicators such as dissolved oxygen, temperature and pH – is better or no worse today than it was a few decades ago. And even if we now recognize that many of our streams have indicators of microbial contamination, we must also recognize that cases of human illness from drinking water and water contact recreation are less today than in the past.

This leads me suggest that there is a value in reflecting on how far we have come in accomplishing our environmental goals. If the Virginia Water Center had access to the resources to fund the work I would commission a natural history (often using oral history) of each watershed and river in the Commonwealth. Increased understanding our environmental history and successes would be an uplifting and informative contribution to our policy debate that at times seems to be stuck in a pessimism about what we have and can accomplish.

Point 2: Pursue Realistic Goals

Most would agree that in the 21st century we will judge environmental quality by a watersheds ability to support aquatic and related terrestrial life – and not just by chemical water quality parameters. Protection and rehabilitation of a watershed's living resources has been called the restoration *of aquatic ecosystems*. Restoration as an organizing idea for watershed management was given a significant boost a few years ago by the NRC Committee on Restoration of aquatic ecosystems. When we prepared that report we defined a restored watershed as one that has been returned to some "pre-disturbance" conditions, related to the chemical quality of the water, the pattern and timing of water flows and the species composition. We debated and never resolved how far back to go in a watersheds history to define "pre-disturbance", but you get the idea.

The committee argued that important management information can be gained by understanding the watershed processes of the pre-human disturbance condition – or at least some historical watershed condition – gleaned from all sources of data and natural and human history. However, we must be careful that the undisturbed – often historical – condition of a watershed without people does not become the standard to be achieved.

Think about bio assessment for watershed water quality management. Critical to bio assessment is the selection of a reference stream. A September 1997 EPA publication states that the reference watershed condition should be "… an unimpaired stream – defined as a condition where human activity is at a minimum." This EPA document goes on – "The overall goal of establishing a reference condition is to describe the natural potential of the water body and habitat types characteristic of the region, independent of the extent of human degradation." Thus we are instructed to choose a reference condition for defining a *healthy watershed as watershed where there are few people or limited human activity*.

For assessment science this may be the reasonable reference condition and we might learn much from studying pristine waters, but the pre-human disturbance condition in any watershed does not establish an environmental goal – a standard (if you will) -- to be met. To seek to achieve the reference or historic condition in all watersheds is unrealistic. However, many have argued – and that was an NRC committee argument with which I had sharp disagreement - that we should set lofty goals even if they cannot always be achieved. But does it make sense to say a healthy watershed *only* can be one that has had no human influence?

If every stream that fails to match a reference condition is declared impaired, and if each impaired stream must be returned to a reference state, if that restoration requires significant cash outlays and imposes high opportunity costs, then is restoration to a pre disturbance -- or reference condition -- justified? This will be a central analytical question to be addressed in watershed management in the next century, but it is one we are ignoring at the present time especially in out TMDL studies that focus on hydrology and water chemistry more that socio-economic values to be achieved and costs.

Instead maybe we should be more willing to start talking about watershed triage – for degraded systems we should only make sure they are not health hazards. Maybe we need to start talking about sacrificial watersheds so we can direct urbanization and dense development to some watersheds in order to eliminate the sprawl development that threatens all watersheds. We need to think of watershed management as an investor thinks about maximizing returns from a portfolio of investments. The investor's highest returns come from emphasizing only certain instruments in that portfolio. Likewise watershed management must choose which watersheds will be favored. Some streams must support carp so others can support trout!

Point 3: Any Goals Require Tough Programs

A water quality goal means a waste discharge cap. A cap is a constraint that all economic and population growth in a watershed must stay within. And when we look past treatment plants and farms we all, in our daily lives, are pollution sources because we all use the landscape. We all must face a constraint.

Return for a moment to a nitrogen loading problem and imagine a TMDL-like limit on N delivered to a watershed. There are many sources of nutrients: industrial facilities, crop and animal agriculture, point source POTWs – a unit of municipal government, urban storm water and other urban sources. Imagine that all these sources categories had a responsibility to limit their nutrient discharges so that water quality goals were maintained, even in the face of growth.

Lets get specific.

Farmers in all counties must – together – would agree to a county wide nutrient reduction program. The farmers form an association and that association must comply with the limit. Farmers would work cooperatively to find the best way to meet their responsibility for the water quality goal. Farmers could choose to go-it alone, if they wished to be outside the system but would be expected to take certain prescribed actions. Counties could combine their efforts to jointly meet the requirements.

The POTWs and other point sources form an association and the group as a whole could agrees that their combined discharge will be under some limit. It is up to them to decide who controls what amounts. If the association limit is exceeded the association would pay an offset fee to a wetlands restoration and riparian buffer establishment fund.

Each municipality in the watershed is assigned a limit on the N that can leave its jurisdiction and enter the river. The counties and cities could choose how to achieve that reduction and might even partner with other local governments. If the load limit is exceeded the local government might pay a fee to a wetlands restoration and riparian buffer establishment fund.

There might be some funding provided to based on the amount of reduction achieved, but most of the financial responsibility would fall on the polluter. The polluter would pay – and we are all polluters.

This may seem to be pure fiction. But, in fact, this is a bare bones description of the Neuse River Basin Nutrient Sensitive Waters Management Strategy, as adopted by the North Carolina State

Environmental Management Commission in December of 1997. Can it come here? There were unique factors that lead to the Neuse plan (hogs, pfesteria and tourism), but in the future the Neuse model may spread.

Living with a cap means stressing pollution prevention and new waste control technologies. We need to create the financial rewards for such new strategies, but current water quality law does not offer such rewards. This is a more complicated story than I have time for here. Instead, I do want to ask you to think about what pollution prevention means when we talk about land use? I think it means land settlement to minimize and redirect impervious and semi pervious surfaces. It means intensification of development in some places in order to reduce development elsewhere (everyone can not live by a trout stream!). Pollution prevention means tackling the way we settle the landscape and creating and adopting the tools for growth management has not been easy in Virginia or elsewhere.

Point 4: A New Capacity in Models and Monitoring

TMDL models must lead to defining water quality goals and implementing plans that lead to the chosen water quality improvement. The EPA calls for stakeholder involvement, consensus decision making , etc. in the TMDL process. Negotiated solutions will answer tough questions: How many fish are enough? How clean is clean enough? What will it cost? Who pays and how? In these processes models can never define a "best" course of action. Models instead should be in the service of the parties to the negotiation.

I especially like the “shared vision modeling” philosophy for supporting this type of process. In the shared vision process the models are built with the stakeholders and are agreed to by the stakeholders. In this way the models do not appear as a black box and so trust in the model analyses of costs and restoration effects is built from the beginning. If the shared vision approach is not taken, at least the models should be as transparent and intuitive as possible. I am afraid our rush to contract out modeling makes this approach impractical. Shared vision modeling demands increased capacity in the agencies for modeling and working with stakeholders. I am afraid we will only be able to hide behind contractor models for a short while. The models (which are weak) will soon be challenged when the costs of implementation are realized. The TMDL lawsuits and conflicts have just begun!

Models are important, but we still pretend that a textbook-planning model will be possible here: we find problems, define alternative solutions, choose the best alternative to solve the problem (send the bill to Richmond or Washington) and move on. This model of the all-knowing technical expert and the deep pockets where everyone pays for everyone else’s water improvements needs to be replaced by a more humble image of the tentative and unsure resource analyst working with stakeholders who must pay the bills.

For these reasons in the next century resource managers will increasingly invoke the term adaptive management. Adaptive management admits the limitations of current knowledge and data as a guide to decision making. Adaptive management makes knowledge creation an objective for watershed management. Adaptive management is a long-term research program conducted in a real watershed, where the trials are actual restoration actions. Adaptive management demands that restoration actions be taken with some experimental design in mind. Then actions are taken, monitoring occurs, and based on that feedback and the new insights gained, adjustments to the restoration plan are made.

By attempting --and then monitoring -- watershed restoration efforts, the state of the environment and the state of the science can advance together. When measurements have been made there has been little effort to assure that sampling is representative of the entire stream cross-section or the entire watershed. Stream flow data is rarely correlated with water quality sampling. And finally monitoring data have not been collected or interpreted to with the intent of linking those data to ecosystem characteristics. At present the data are just not up to the questions we will be asking.

Improved monitoring will cost money, but it is important to not just ask for more money. In return for financial support agencies responsible for gauging and monitoring should prioritize sites and design sampling stations to answer specific questions about the status and trends. This is the essence of adaptive management and it will help identify the consequences of management actions.

To make this kind of monitoring cost-feasible I look forward to significant breakthroughs in remote sensing, imagery, instrumentation, testing and new methods of continuos data transmission and analysis. In a decade our capabilities in monitoring will be significantly advanced. Perhaps the confidence bands that can be attained with the new technologies will be wide, and as a result, the results may be questionable for court cases or regulatory actions. However, we must understand that not all data need to be of the quality and defensibility to win a court case.

In closing

Have perspective while we pursue realistic goals, but remember these goals will require tough programs that will demand a sound science base and a new capacity in models and monitoring.

3. Assessing Violations of Water Quality Standards Under Section 303 (d) of the Clean Water Act

Eric P. Smith, Keying Ye, Chris Hughes and Leonard Shabman

Abstract

Section 303(d) of the Clean Water Act requires states to identify water segments where loads of pollutants are violating numeric water quality standards. The Environmental Protection Agency (EPA) guidelines require a stream segment to be listed as impaired when greater than 10% of the measurements of water quality conditions exceed numeric criteria. This can be termed a "raw score" assessment approach.

A comparison of the raw score method with alternative statistical procedures found that the raw score approach is prone to Type I error (a false declaration that standards are violated) while the binomial statistical method is prone to Type II error (a false declaration of no impairment). A Bayesian approach and acceptance sampling by variables were also evaluated.

An inference that water quality is impaired initiates the TMDL planning requirement. Falsely concluding that a water segment is impaired results in unnecessary TMDL planning and pollution control implementation costs. On the other hand, falsely concluding that a segment is not impaired may pose a risk to human health or to the services of the aquatic environment. An approach that recognizes Type I and II error in the TMDL assessment listing process is suggested.

Keywords: TMDL, Monitoring, Standards, Assessment, Binomial, Bayesian methods

INTRODUCTION

The Total Maximum Daily Load (TMDL) process now dominates water quality policy discussions. Policy reviews (Houck, 1999), lawsuits (USEPA, 2000), proposed regulations (USEPA, 2000a) and congressional interest (GAO, 2000; US Congress, 2000) all have been directed to what had, until recently, been an obscure provision of the Clean Water Act. The TMDL process originates with Section 303(d) of the Clean Water Act. That section requires states to conduct an assessment of, and then report on, the condition of their waters. In practice this means that the states review the water quality conditions in specific segments (a lake, bay or river) "upstream" of water quality monitoring locations.

Each state's 303 (d) impaired waters list identifies segments where anthroprogenic loads of pollutants are leading to violation of water quality standards. The listed segments must remain on the list until the pollution problem has been addressed or until subsequent monitoring or other information suggests that the segment was misclassified. Addressing a water quality problem for a 303 (d) listed water is a complicated and potentially expensive process. First, a watershed study is initiated to establish the maximum quantity of each pollutant that can be discharged to a segment if the segment is to meet water quality standards. Once the maximum load is defined there are a series of steps to allocate responsibility for load reduction to identified pollution sources and to secure those reductions over time. These steps constitute the TMDL watershed study and implementation plan. (REF)

The TMDL planning step for each listed water segment can be costly. In comments to the EPA the states agencies concluded that 25% of TMDLs will be simple and will cost $50,000-$200,000, 65% of TMDLs will be of moderate difficulty and will cost $300,000-$400,000, and 10% of TMDLs will be complex and will cost $600,000-$1,000,000 (US Congress, 2000). A state may have hundreds of segments on its impaired waters list. (USEPA, 2000b). Implementation of a TMDL plan imposes additional and perhaps substantial pollution control costs.

Because of the significant costs for addressing pollution of waters on the impaired waters list, it is appropriate to review how the list of impaired waters is constructed. Such a review is especially warranted because water quality standards, monitoring protocols and guidelines for assessing data were developed before the TMDL program took on its current significance and may have been developed for different purposes.

A review of the 303(d) listing process might examine the basis and intended purpose of the water quality standards themselves. Also, such a review might evaluate the monitoring protocols that secure the data used to make the listing determination. In this paper, we review the guidelines for interpreting the monitoring data that are collected. Specifically, in this paper we review the EPA assessment guidelines for comparing sample measurements of water quality conditions with numeric ambient water quality standards.

Numeric water quality standards are measurable criteria for dissolved oxygen, temperature, pH and fecal coliform bacteria counts. Critical to the 303 (d) assessment is the monitoring data collected by a state's environmental department to assess whether stream conditions meet standards. Cost realities, given the need for statewide monitoring and the fact that most monitoring is for enforcement of point source discharge permits, results in limited stations and limited number of samples for each station. For example, Virginia waters are among the most monitored in the nation with over 17,000 miles of monitored waterways. Virginia's significant monitoring program collects data at each station on a quarterly basis. The 303 (d) assessment occurs every two years so the 303 (d) assessment might be based on two years of data at a particular station (approximately 8 observations.) (VWRRC, 2000). The reality of limited data must be recognized in the 303 (d) assessment process.

The assessment challenge is to interpret the limited amount of sample data to determine whether the stream is violating standards, recognizing that the samples are affected by variability in human activity and natural or background conditions. It appears that the EPA guidelines recognize this variability because the guidelines do not require a water to be listed if less than 10% of the sample measurements do not violate the standard (USEPA, 1997). In effect, the assessment guidelines imply that a violation of the numeric criterion is acceptable 10 % of the time. This willingness to accept failure to meet a numeric criterion might reflect a policy judgement on the “tolerable” level of degradation. However, it also might reflect a recognition that there is variability in natural background conditions and in human activity. In addition, measurement errors in the analysis of the samples collected could be yet another reason why the numeric standard might be violated in a sample.

The EPA guidelines suggest what can be called a “raw score” approach to data interpretation. Because the number of samples are small (as few as 8 in an assessment period) the approach requires truncation. If there are 5 samples and one or more exceed the standard, the site is declared impaired. The same is true for all sample sizes between 1 and 9. For sample sizes between 10 and 19, one sample is allowed to exceed the standard but not more. However, the raw score approach does not allow the water quality assessor to consider the likelihood and costs of making an erroneous listing. Instead of asking whether less than 10% *of the samples violate the standard*, the alternative is to ask whether the segment meets water quality standards 90% of the time. A statistical representation of this different perspective is shown in Figure 1. In Figure 1 the measurement is a concentration of some contaminant in the ambient water. The distribution of the water quality parameter may be drawn to represent the likelihood of ranges of values. As displayed, the water quality standard requires that a concentration of 2.5 should be met 90% of the time although some measurements may exceed the standard naturally.

Suppose 8 samples are taken and a raw score analysis is completed. If one of the samples (10%) exceed the standard the site would be declared impaired. However, the one sample that violates the standard might be attributed to natural variability or an unusual human activity. In this case the site may be classified as impaired when in fact this is not the case. This error is referred to as a Type I error. Another error may occur when a site is truly impaired but the sampled measurements from the site do not exceed the standard and the site is not declared impaired. This error is referred to as a Type II error.

In this paper, the error rates associated with the raw score approach and related approaches are evaluated. The comparisons are made in terms of Type I and Type II error rates. One alternative to the raw score approach is a statistical approach based on a Binomial test for violation. Both the raw score and the Binomial methods treat the sample observations as binary values, either exceeding the standard or not exceeding the standard. We describe two other methods that use additional information. A Bayesian version of the Binomial test uses prior information about violation probability. Acceptance sampling by variables is a method used in quality control that changes the problem of exceeding a standard to one of exceeding a mean. These methods are evaluated in terms of the error rates and compared with the raw score and Binomial method. This evaluation of alternative approaches leads to a recommendation for improving water quality assessments in the 303 (d) process.

Statistical Approaches

The null hypothesis is that the site is not impaired. The alternative hypothesis is that the site is impaired. Therefore, the segment may be declared impaired when it is not (Type I error or a false positive) or the segment may not be designated as impaired when in fact it is (Type II error or false negative). The error rates are bounded between 0 and 1, with 0 indicating no error. However, given the sample sizes likely to be available, either type of error will not closely approach zero.

Because both Type I and II errors always will be present water quality managers must choose the type of error they wish to minimize. In principle, this choice should be based on an explicit consideration of the consequences (costs) of being wrong. Costs may be financial outlays made by governments or private individuals. Costs might be forgone public values that may not be reflected in markets. In the following sections the tradeoff among error types is considered without regard to the cost of being wrong. Cost considerations are discussed in more detail in the concluding section.

The errors associated with different tests were simulated and then compared with the raw score approach. The raw score approach uses limited information to make the impairment determination. An alternative to the raw score, the Binomial testing approach, uses the same limited information (. Both the raw score and bionomial test approaches only consider whether the monitored data exceed or do not exceed the numeric criterion. Acceptance sampling by variables considers the magnitude of the deviation of the measurement from the standard. One way to increase the information for decision making is to increase sample size but the Bayesian approach draws information from other sources about the probability of violation. Two questions are of interest. The first question is whether any of the statistical approaches would reduce either type of error. The second question is whether they can be practically applied in the assessment process.

To compare the error rates, the acceptable probability of violation is set at 10%. The analysis assumes that the water quality parameter of interest has a distribution that does not change over time and that the samples collected are independent of each other.
Based on these assumptions the probability a given sample exceeds the standard may be modeled as a random variable. The listing decision process may be viewed as a test of the null hypothesis that the probability of violation is less than or equal to 10% versus the alternative that it is greater than 10%. The Type I error rate may then be computed. To compute a Type II error rate for this illustration (given the site is impaired, how likely is it that we do not detect impairment) the true probability of exceeding the standard impaired must be specified; this percentage is set at 25%. Using this framework, the distribution may be used to calculate the error rate for the raw score method by calculating the probability of not rejecting the null hypothesis (i.e. getting fewer than a significant number of violations).

The Binomial Method

When applying the binomial approach, observations exceeding the numeric criterion are assigned the value 1 and those that do not are assigned the value 0. Then if n samples are collected the number of observations exceeding the criterion (number of 1's) may be viewed as a Binomial random variable with parameters *p* and n. The parameter *p* is interpreted as the true but unknown probability of exceeding the standard. Using the Binomial model, one may then test the hypothesis that $p<0.10$ (not impaired) versus the alternative $p>0.10$ (impaired). These hypotheses may be viewed as a null hypothesis that the site is not impaired versus the alternative that the site is impaired. With this approach error rates associated with impairment declarations may be evaluated and a process to limit the error rates can be described.

In a typical statistical analysis the Type I error rate is chosen by the assessor, perhaps in consideration of costs of being wrong. If the rate chosen is .10, then there is a 10% chance of making a Type I error. For the Binomial method, the choice of the Type I error rate determines the "cutoff" value. For a given sample size n, the cutoff is selected as the number of violations to make the probability of this many or fewer violations to be as large as possible but less than the Type I error rate, assuming the null hypothesis of no impairment is true. Given the cutoff and the alternative for the frequency of violation, the Type II error rate for sample size n can then be calculated. The Type II error rate may be reduced by choosing a lower Type I error rate (for example .05) and/ or by increasing sample size.

The Bayesian approach

In the above analysis, the probability of exceeding the standard is treated as fixed and the data (i.e. does the sample exceed the standard) is treated as random. A Bayesian approach (Berger, 1985) computes the probability that the site exceeds the standard by treating the proportion exceeding the standard as a random variable that has an associated distribution. Initially the form of this distribution is based on previous information and is referred to as the prior distribution. After data are collected, the prior is updated and the data and prior are used to compute the posterior distribution of the exceedence probability. Based on this posterior distribution, a decision may be made using either a cut-off approach or a ratio approach (Bayes Factor). This process is described in more detail in Appendix 1.

Suppose there is a Binomial random variable with associated sample size n and parameter *p* (i.e. $X \sim bin(n,p)$. This is whether test $H_0 : p$? p_0 versus $H_1 : p$? p_0, where p_0 is a constant between *0* and *1* (in the current problem, it is 0.10). Suppose now that a prior distribution of *p*, ? (p), can be specified.

A prior distribution for *p* might be developing by introducing new information to the analysis. One possibility is to use samples from other sites that are not impaired. For the unimpaired sites information would be collected and the probability of exceeding the standard calculated.

Given observations and a prior distribution, Bayesian criteria can be used to make an inference about *p*. Using the prior and data, the posterior distribution of *p* may be written as

$$? \, (p \,|\, x) \ ? \ \frac{f(x \,|\, p) ? \, (p)}{?_0^1 f(x \,|\, p) ? \, (p) dp}.$$

This new distribution represents current knowledge about the probability of a violation found by updating the prior information based on the data from another site. Using the above distribution, the posterior probability of the null and alternative hypotheses may be calculated. For the null hypothesis H_0, that the site is not exceeding standards, the probability is computed as $?_0 \ ? \ P(H_0 \,|\, data) \ ? \ P(p \ ? \ p_0 \,|\, x)$. For the alternative H_1, that the site is exceeding standards, the prior may be calculated as $?_1 \ ? \ P(H_1 \,|\, data) \ ? \ P(p \ ? \ p_0 \,|\, x)$. Two approaches for evaluating these probabilities and making decisions are the cut-off method and the ratio method.

Cutoff Method: Use the posterior probability to determine the rejection rule.

Predetermine a probability *q* (analogous to the Binomial method Type I error rate, *q* might be specified as *0.10*). If the posterior probability that the alternative hypothesis is true exceeds *q*, then we reject the null hypothesis and conclude that the water is impaired i.e. $P(H_1 \,|\, data) \ ? \ q$. The quantity *q* is referred to as the posterior cutoff.

Ratio Method: Use the Bayes factor to determine the rejection rule.

The Bayes factor of H_1 against H_0 is the odds ratio of the posterior probability of H_1 against H_0 divided by the odds ratio of the prior probability of H_1 against H_0. It can be expressed as

$$B_{10} \ ? \ \frac{P(H_1 \,|\, x)}{P(H_0 \,|\, x)} / \frac{P(H_1)}{P(H_0)}.$$

A large value of the Bayes Factor would indicate that the null hypothesis is not correct. Kass and Raftery (1995) (see also Jeffreys, 1961) suggest that when B_{10} is between 3 and 20, the evidence of H_1 against H_0 is strong. Bayes Factor cutoffs of 3 and 10 were used in our examples.

The difference between the cutoff and ratio methods is in the importance given to the prior. The influence of the prior is usually diminished if the Bayes factor method is used. Because of the possible subjectivity of the prior, decision makers may want to choose to use the Bayes Factor approach. If the available prior information is empirical the cutoff method might be adopted.

Both methods require evaluation of the prior probability of the null and alternative hypotheses. The approach may be extended by using a weighting factor *v* (between 0 and 1) that balances the prior distribution between the null and alternative hypotheses. A value of *v* that is near 1 would indicate a stronger belief in the null hypothesis. In the figures comparing the methods, we refer to this value as $p(H_0)$ or prior(H_0). Details of the computations are given in Appendix 1.

Acceptance Sampling by Variables

The Bionomial and Bayesian methods treat the sample observations as binary values, either exceeding the standard or not exceeding the standard. A criticism of this approach is that the actual values are not taken into account. If the standard is 1.0, measurements of 1.1 and 10 are given equal importance and both are treated simply as exceeding the standard. In quality control one approach that uses all the data to make a decision is "acceptance sampling by variables" (Duncan, 1974).

Acceptance sampling by variables transforms questions about the proportion of samples that exceed a standard into questions about the center (or another parameter) of a continuous distribution. The actual values of the observed variable are used in the analysis procedure. The method (as typically applied) assumes that the measurement of interest has a distribution type that is known. The standard is used to estimate the parameter of interest, treating all other parameters as known.

Let *Y* be the water quality variable being measured. Suppose *Y* has an associated standard upper specification limit *U* such that an observed value of *Y* that is greater that *U* is deemed unacceptable. A particular value of *Y* could simply be classified as above *U* (does not meet standards) or below *U* (meets standards).

In the case of an assumed normal distribution for the measurement, the variance $(?^2)$ would be assumed to be known. Then the mean may be calculated based on the standard and the known standard deviation. If p_0 is the proportion associated with the standard (in our case p=0.10) then the value of the mean would be calculated as

$$?_0 \; ? \; U \; ? \; ? \, z_{1 ? p_0.}$$

where z represents the value from the standard normal distribution associated with a cumulative probability of $1\text{-}p_0$. Error rates enter in a similar manner to the Binomial test. The hypothesis becomes H_0: ? ⁇ ⁇ $_0$ and the alternative is H_1: ? ? $?_0$. The Type I error rate for the test would be specified by the user. The Type II error rate is calculated by choosing an appropriate value of the mean given the alternate hypothesis is true. This may be calculated in terms of the exceedance probability as $?_1 \; ? \; U \; ? \; ? \, z_{1 ? p_1.}$

RESULTS

First the Type I error rates for the raw score approach called for in the EPA guidelines and Binomial approaches are compared. Then the error rates for the Bayesian and acceptance sampling methods are compared with the Binomial and raw score methods. These alternative methods require additional information.

Binomial method

The Type I error rates are compared using calculations of Binomial probabilities under different sample size scenarios where *p* was set to 0.10. The probability that a site is declared as impaired when in fact it is not (false positive) is displayed in Figure 2. Note that the figures are jagged, with each spike corresponding to a change in the critical value (i.e. number of violations required to declare impairment). The Binomial method controls for Type I error (i.e. it is always less than or equal to a preset value of 0.10 or 0.20) and the raw score approach does not. With the Binomial method, the Type I error rate is fixed at some value (?) that is an upper bound on the error. The actual error rate for the Binomial method is determined by computing the probability of getting less than "x" samples exceeding the standard. The greatest probability that does not exceed ? ? is the actual error rate. Figure 2 shows that the Type I error rate (a false declaration of impairment) for raw score method is quite high relative to the Binomial. For example with a sample size of 10 the Type I error rate for the raw score approach is 60%. With one more sample it drops to 30% (an example of the effect of truncation), but is still 3 times the Type I error rate of the Binomial approach. The Type I error rate then rises rapidly with sample size until the cutoff changes. As sample size increases the Type I error rates for the different methods do not converge. Thus, relative to the Binomial approach the raw score approach is prone to Type I error (a false declaration of impairment). The tendency of the raw score approach to lead to Type I error *is not* mitigated by increased sample size.

Figure 3 presents Type II error rates. We assume for the computations that the segment violates the numeric criterion 25% of the time and so the segment violates standards, however the violation is not detected. In statistical terms this represents failure to reject the hypothesis that the violation rate is equal to 0.10 when in fact the violation probability is 0.25. In this case, Figure 3 is reversed from Figure 2. The Binomial method is prone to Type II error relative to the raw score method. For example with a sample size of 10 the Type II error rate for the Binomial is about 5 times the rate for the raw score approach. With one more sample the difference drops to 2 times (a result of the effect of truncation). As sample sizes get larger the Type II error rates do converge to zero, which is to be expected. These results are appropriate for the case of a critical error being associated with a violation probability of 0.25 and a preset Type I error rate of 0.10. Figure 3 also shows the result of increasing the acceptable Type I error rate to 0.20. The results indicate that the chance of a Type II error using the Binomial method decreases with an increase in the Type I error rate and with increased sample sizes. For sample sizes of n=8, the Type II error is 0.37 for a Type I error of 0.20 while for a Type I error of 0.10, the Type II error is 0.68. For n=20, the error rates are 0.23 versus 0.41.

One possible approach to addressing the different error rates is to seek to make Type I and Type II error rates the same for each sample size (Bross, 1985). In effect, this implies that the cost of Type I and Type II errors are the same. In Figure 4, the error rates are plotted using a Binomial test with the null *p*=0.1 and the alternate *p*=0.25 with cutoff values chosen to make the error rates as close as possible. If there are at least these numbers of samples exceeding the standard, the site is declared impaired. Cutoff values are plotted on a second vertical axis. Note that for small sample sizes it is difficult to equate the error rates although there are sample sizes where the error rate lines cross. These are associated with different cutoffs. Examples are n=10, Type I error=0.26, Type II error=0.24 and cutoff=2; n=16, Type I error=0.21, Type II=0.20, cutoff=3; n=22, Type I error=0.17, Type II error=0.16, cutoff=4. Note that if it is desired to have both error rates around 10% then a sample of size 34 would be required (cutoff=6, Type I error=0.12, Type II error=0.11). With sample sizes currently used, it is difficult to obtain error rates below 20% for both types of errors.

Relative to the EPA raw score approach, the Binomial method is more prone to Type II error. The tendency toward Type II errors in either approach *is* mitigated by increased sample size, although even at sample sizes over 20, Type II error rates for the Binomial are around 2-3 times higher that the raw score approach. An advantage of the Binomial approach is that for sample sizes (around 20-25), both error rates may be reduced under the Binomial method. Specifically, at sample sizes of around 25 Type I and Type II error rates with the Binomial method can be made around 20% for each type of error.

The Bayesian method

To evaluate decision rules based on the Bayesian method, we considered three situations for method 1 with a uniform prior for *p* (*v=0.75, 0.90* and *0.99*, *q=0.1*) and two values of cutoff for method 2 (using Bayes Factors of 3 and *10*). Also, for comparisons, we include the decision rules using the Binomial method and the EPA method. We computed the Type I and Type II error probabilities of the decision rules considered above. The Type I error rates are displayed in Figure 6. The interesting feature of this plot is that as sample size increases, the error rate for the raw score method remains quite large. Thus the raw score method will tend Type I error. Type II probabilities are displayed in Figure 7. Note that as expected, as the number of samples increases, the Type II error rate decreases for all methods. Note that Figure 7 indicates that the raw score method has the smallest Type II error rate.

Figure 5 shows the number of violations required to declare a site as impaired for the different rules. The Bayes factor method with *q=10* and method 1 using *v=0.99* are most prone to Type I error and the raw score approach is the most prone to Type II error. Method 1 using *v=0.75* is also prone to Type II error (notice that *v=0.75* gives 25% of prior information that H_1 is true). The Bayes factor rule with *q=3*, Binomial rule and method 1 using *v=0.90* are all in the middle and are close to each other. With a Bayesian approach there is strong dependence on the prior information.

When we try to balance both error probabilities (Figure 8) gives interesting results (for the sake of simplicity in graphing, we do not include method 1 using *v=0.75*, the Bayes factor method with *q=0.10* and method 1 with *v=0.99*). In this graph, the averages of the two types of error probabilities are drawn. It is obvious that the raw score method can not balance the two error probabilities, especially for larger sample sizes. Also, averages of the averages for each decision rule are calculated for *n=2* to *n=100* and for *n=8* to *n=24* (most likely range for the monitoring problem). Method 1 (posterior cutoff of 0.1 and prior H_0 of 0.75) is a method that compromises between the Binomial and raw methods. The results suggest that if sufficient information is available and the proper posterior cutoff selected, then the Bayesian method may improve the decision process relative to the Binomial and raw score methods.

The Bayesian approach is a very flexible approach. By controlling various parameters, it is possible to obtain tests with good properties when the sample size is reasonable (around 20). With small sample sizes, it is difficult to make the combined error small. In applying the method, the difficulty is obtaining the prior information. Because the weighting factor and prior probability of acceptance can greatly change the error rates, it is important to have good estimates of this information. If the approach is to be implemented then studies and models are required so that priors may be calculated.

Acceptance Sampling by Variables

Error rates associated with Acceptance Sampling by Variables are presented in Figure 9. Type I error rates are exact and may be specified at a particular level by the agency. By selecting the z value from the normal distribution with a specified probability we are selecting the Type I error rate. Type II error rates are presented in Figure 9 for the case of variance set to 1.0. The Type II error rates for this method are superior to that of the Binomial method and match those of the raw score method once sample sizes exceed 20. In this example, the variance was chosen to make the error rate small and larger variances will have increased error rates. The figure suggests that if the variance is known or can be well approximated, significant gains might be achieved by using the actual observations.

Acceptance sampling by variables has several important assumptions. First is the requirement that the distribution is known. Typically this would be a normal distribution or lognormal distribution. In some instances, it is possible to assess the distribution type to be used. For example, if sufficient data is available at a site a test of distribution type can be made and evaluated. Also, data from other sites that are similar may be combined to give an overall assessment of distribution type.

Second, the standard is used to estimate a single parameter. If the distribution has more than one parameter, then other parameters are assumed to be known. In the case of the normal distribution, the standard is used to estimate the mean and the variance is treated as known. The case is similar for other distributions such as the lognormal distribution. The method may be extended to the situation where the variance is treated as unknown. In this case however, the Type II error rates should increase due to additional uncertainty due to having to estimate the variance.

Acceptance sampling by variables is a promising method. The method requires the variance to be known, but this may not be a barrier if there is information for calculating this information from sites that are not impaired.

Conclusions and Recommendations

State listing of impaired waters under Section 303(d) triggers what is now called the TMDL process. Once on the impaired waters list, significant pollution control planning and implementation costs can be incurred. On the other hand, failure to identify and address significant water quality impairments is not acceptable public policy. For these reasons, the TMDL process has become a source of intense debate among EPA, the states and organizations representing environmental groups, waste dischargers and non-point sources.

Because of the significant consequences of a Section 303(d) listing, or of a failure to list, the interpretation of sample data is especially critical. The EPA guidelines call for the use of a raw score approach to data analysis. However, in other contexts, raw score approach has been criticized for a number of reasons, including error rates (Barnett and O'Hagan, 1997). The calculation of the likelihood of Type I and Type II error under different methods is a critical step in selecting a preferred approach and for applying that approach for interpretation of sample data. Clearly, when choosing an approach, the Type I error rate and the sample size, the assessor is implicitly or explicitly making a determination of the relative acceptability of Type I versus Type II error. This choice about acceptable error has been called a risk management policy decision, but is now ignored in the Section 303(d) listing process.

An advocate for the statistically based methods is *implying* that it is more important to avoid a Type I error (incorrect declaration of impairment) than a Type II error. An advocate for the raw score approach is implying that it is more important to avoid a Type II error than a Type I error. The Binomial method and other statistical approaches can be applied to address the balancing of error rates while the raw score method remains prone to Type I error. However, concern for Type II error is important and if a statistical approach is taken assessors may be more concerned about that type of error.

Ideally, risk management policy decisions would be made based on explicit consideration of the consequences (costs) of being wrong when making a declaration of impairment versus no impairment. In the Section 303 (d) listing process consequences might be thought of in terms of "costs" that follow from the impairment determination. Cost may be financial outlays made by governments or private individuals. For planning and pollution control, costs also might be forgone public values that may not be reflected in markets as people avoid use of waters that are listed as impaired and calculation of these costs may be more or less certain. Because many impairments are for violation of the fecal contamination standard the following discussion is for that standard.

Consider first the cost of a false positive (Type I) error. At present, if a stream is declared impaired a TMDL process will be initiated. There is a cost to TMDL planning and modeling that is significant financial outlay. Each study is a claim on a limited agency budget and so available resources are spread out more thinly as the number of segments listed as impaired increases. Therefore, in the face of limited budgets, a segment that is declared impaired when it is not impaired may divert limited resources from actual to false problems. Once the impairment is declared there may be public avoidance of the segment and a loss of public use values. Once again, if the segment is not impaired then those values forgone are an unnecessary cost. Next, planning moves forward and there are implementation costs (BMPs, etc.) imposed to change practices at the suspected source of the pollutant. Such implementation costs might be imposed on public agencies and the private sector at the end of the TMDL process. The preceding logic helps explain the desire to avoid a Type I error.

Now consider the possibility of Type II error where the pollutant be one that has the potential to impose a cost in terms of human sickness. If human health were at issue the water quality assessor might wish to avoid Type II errors. If the problem was, for example, low DO problem in a stream and fish habitat the cost of a Type II error might be quite different. On the other hand, in the case of microbial contamination there is much uncertainty about the source and pathways for the pollutant and the effects on human health. (Hagedorn, 2000) There may be uncertainty about whether the measure contaminant poses a health risk, there may be uncertainty about the exposure to the pollutant (who swims in a creek and when for example), there may be uncertainty about whether the exposed population will in fact be affected by the contaminant even if it is in the segment, and finally the severity of the reaction to the exposure may be uncertain. None of the preceding is to say that the costs of illness are to be ignored, but there is much uncertainty about the costs even if the pollutant is present.

On balance the cost of Type I error is more certain and more immediate, but there are no general data or calculations that would allow for a cost balancing of error types. However, social attitudes toward environmental risk suggest that we want to be risk averse and would tend to avoid Type II error even if Type I error is more likely and has a more certain cost.

With this cost uncertainty in mind and because inferences about water quality conditions now are made using small samples, there is no compelling reason to uniformly favor one approach over another. For small samples (less than 12) the raw score approach might be used along with the Binomial method (as described below). However, if sample sizes are increased (around 20 – 25) the assessment process should rely on statistical procedures because of ability to manage and measure Type I and Type II error. In fact, the raw score approach in insensitive to sample size increase and so is not made more powerful with more data.

Alternative methods such as a Bayesian approach or Acceptance Sampling by variables provide potentially useful alternatives but require additional information. The Bayesian method may result in better testing if information regarding the probability of declaring a stream impaired based on prior data. Acceptance sampling provides a useful alternative when there is information about the variance of the chemistry measurements for unimpaired streams. Acceptance sampling by variables should be considered if there is a significant variance in the measured data that can not be easily explained by background conditions such as differences in flows.

It has also been recognized that Type II errors are more likely to occur with the statistical methods than with the raw score approach. While the increase sample size will reduce the probability of Type II error, water quality assessors may feel that the statistical approaches are still too prone to Type II error. One strategy for reducing the Type II error would be to increase the type I error rate.

Given the information routinely used in an assessment, the Binomial method should replace the raw score approach when sample sizes are larger than those now being used. However, the concern over Type II error (declaring an impaired stream to be unimpaired) suggests a strategy for interpreting small sample sizes. With larger samples, in order to be sensitive to concerns over Type II error rates, agencies should be encouraged and provided the resources to increase sample sizes for the assessment process.

Given the findings and implications of this review a sampling logic for the assessment listing process is suggested in Table 1.

Table 1. Logic table for assessing site impairment.

ASSESSMENT STEP	LOGIC
1. Is the sample size "small" (~=10)? If yes go to 2; If no go to 5.	Current sample sizes are small. In the longer term sample sizes should be larger before there is full reliance in statistical measures.
2. Does the Binomial analysis with small samples indicate impairment? If yes, then list as impaired If no, go to 3.	A site that fails the Binomial test (reject the hypothesis that the site is not impaired) should be listed as impaired on the 303 (d) because of the low probability of Type I error, for any sample size.
3. Does the raw score with small samples indicate impairment? If no, then do not list as impaired If yes, list as threatened and go to 4.	A site that is not impaired using the raw score test *with small samples* should be considered unimpaired, because of the raw score approach tendency to Type I error. However because the raw score test is highly prone to Type I error and because of the high cost of Type I error, failure to pass the raw score test should be interpreted as signaling a *possibility* that standards are being violated.
4. Design and implement a monitoring plan to increase sample size to at least 20-25 samples for the desired period of analysis. Proceed to 5.	The failure to pass the raw score test with small samples should lead to more intensive study of the sites (more sample collection), rather than an impairment determination. The 303 (d) assessment list could include a category of waters that are under more intense monitoring. Action may involve sampling with remote technology; more frequent sampling with traditional approaches or use longer record adjusted for trend to increase the sample size. (Zipper, et. al. 1998)
5. Use the Binomial analysis of the sample data to make listing decisions, with alternative Type I error rates. If variability in sample measures is high use acceptance by variables as well.	

References

Barnett, V. and A, O'Hagan. 1997. Setting Environmental Standards. Chapman and Hall, London.

Berger, J. 1985. Statistical Decision Theory and Bayesian Analysis (2nd edition). Springer-Verlag, New York.

Bross, I.D. 1985. Why proof of safety is much more difficult than proof of hazard. Biometrics, 41: 785-793.

Duncan, A.J. 1974. Quality Control and Industrial Statistics. Irwin, New York.

Hagedorn, C. 2000. "Tracking the Wild-and-Domestic Bacteria." Virginia Water Central, (October), pps. 7-10. available at http://www.vwrcc.vt.edu/pdf/vwc1099.pdf

Houck, Oliver A. 1999. The Clean Water Act TMDL Program: Law Policy and Implementation. Washington D.C. Environmental Law Institute

Jeffreys, H. 1961. Theory of Probability. Oxford University Press, Oxford.

Kass, Robert E.;Raftery, Adrian E. 1995 Bayes factors. Journal of the American Statistical Association 90:773- 795

US Environmental Protection Agency (USEPA). Office of Water. 1997. Guidelines for Preparation of the Comprehensive State Water Quality Assessments. EPA-841-B-97-002B, September, 1997. p. 3-17.

US Environmental Protection Agency (USEPA). Office of Water. 2000. Total Maximum Daily Load Program: TMDL Litigation by State. http://www.epa.gov/owow/tmdl/lawsuit1.html (accessed March 16, 2000)

US Environmental Protection Agency (USEPA). Office of Water. 2000a. Total Maximum Daily Load Program: Proposed Rules. http://www.epa.gov/OWOW/TMDL/proprule.html (accessed March 16, 2000)

US Environmental Protection Agency (USEPA). Office of Water. 2000b. Total Maximum Daily Load Program: National Overview. http://www.epa.gov/OWOW/TMDL/status.html (accessed March 16, 2000)

US Environmental Protection Agency (USEPA). Office of Water. 2000. Total Maximum Daily Load Program: Major Pollutants Causing Impairment by State. http://www.epa.gov/OWOW/TMDL/303dcaus.html (accessed March 17, 2000)

Zar, J.H. 1996 Biostatistical Analysis. 3rd edition. Prentice Hall, Upper Saddle

United States General Accounting Office (GAO). 2000. Water Quality: Identification and Remediation of Polluted waters Impeded by Data Gaps. GAO/T-RCED-00-88. Washington D.C.

United States Congress (US Congress). February 10 and 15, 2000. Subcommittee on Water Resources & Environment Hearing on EPA's Proposed Regulations Regarding Total Maximum Daily Loads (TMDLs), the National Pollutant Discharge Elimination System (NPDES), and the Federal Anti-Degradation Policy. http://www.house.gov/transportation/ctisub5.html and http://www.house.gov/transportation/ctisub5.html (accessed March 16, 2000)

Zipper, C., G. Holtzman, P. Darken, P. Thomas, J. Gildea, and T. Younos. 1998. Long-Term Water Quality Trends In Virginia's Waterways. VWRRC Report SR11-1998. Blacksburg VA. Virginia Water Resources Research Center.

Appendix 1. Details for the Bayesian approach

Two cases are considered below. First we assume that the prior distribution on p is a mix of two distributions, one associated with the null and the second with the alternative. A weighting or mixing parameter is used to combine the distributions together. Calculations can be made compute the posterior probabilities of the null and alternative hypotheses or the Bayes factor. In the second case, the priors are assumed to be uniform distributions on the interval 0 to 1. This suggests less knowledge of the probability of the hypotheses. Details on how to calculate the posterior probability and the Bayes factor are then given below.

(a) Mixture prior: Suppose that $?(p) ? v?_0(p) ? (1 ? v)?_1(p)$ is the prior distribution where *v* is a number between *0* and *1*, and $?_i(p)$ is the density only defined on the region of H_i, for *i=0,1*. The constant *v* can be viewed as the prior probability that H_0 is true. The posterior probability of H_1 can be calculated by

$$?_1 ? \frac{(1 ? v)?_{p_0}^{1} f(x|p)?_1(p)dp}{m(x)}$$

and the posterior probability of H_0, is

$$?_0 ? 1 ? ?_1 ? \frac{v?_0^{p_0} f(x|p)?_0(p)dp}{m(x)}.$$

Therefore, to use Method 1, we can determine the cutoff value *q* to find the decision rule. On the other hand, the Bayes factor can be determined by

$$B_{10} ? \frac{?_1}{?_0} / \frac{(1 ? v)}{v} ? \frac{?_{p_0}^{1} f(x|p)?_1(p)dp}{?_0^{p_0} f(x|p)?_0(p)dp},$$

which no longer depends on *v*.

(b) Uniform prior: Suppose that both $?_i(p)$, *i=0,1* are uniform distributed on the regions. Then $?_1$ can be simplified as

$$?_1 ? \frac{\frac{1 ? v}{1 ? p_0} P(p ? p_0 | x)}{\frac{1 ? v}{1 ? p_0} P(p ? p_0 | x) ? \frac{v}{p_0} P(p ? p_0 | x)},$$

where the probabilities *P(A|x)* are calculated using a *Beta(x+1,n-x+1)* density.

(i) To apply method 1, we need to find *x* such that

$$P(p ? p_0 | x) ? \frac{1}{1 ? \frac{(1 ? p_0)vq}{p_0(1 ? v)(1 ? q)}}.$$

(ii) To apply method 2, if the cutoff is *q*, we need to find *x* such that

$$P(p ? p_0 | x) ? \frac{p_0}{p_0 ? (1 ? p_0)q}.$$

Appendix 2. Details for Acceptance Sampling by Variables

Suppose Y has a normal distribution with mean ? and variance $?^{2}$. The proportion of samples that will exceed U is

$$p ? P(Y ? U) ? P ? \frac{Y ? ?}{?} ? U ? ? 1 ? ? ? \frac{Y - ?}{?} ?.$$

It is possible that some samples will exceed the limit U even if the waterway is not significantly impaired. Let p_0 be an acceptable proportion of samples exceeding U. If we assume ? is constant, p is a function only of ?. Let $?_0$ represent the value of ? that corresponds to p_0 and $z_?$ be the ? quantile of the standard normal distribution, i.e. P(Z? $z_?$) = ?. Then $?_0$ is given by

$$?_0 ? U ? ? z_{1? p_0.}$$

Let p_1 be the minimum unacceptable proportion of non-complying samples and let $?_1$ be the corresponding value of ?. Then

$$?_1 ? \mathrm{U} ? ? z_{1? p_1.}$$

We will normally take $p_0 = 0.10$ and $p_1 = 0.25$. The corresponding values of $?_0$ and $?_1$ can be obtained from the above equations.

For a given sample size of n observations, the stream should be declared impaired if $\bar{y} ? c$ where c is a constant. The Type I error rate p_0 (i.e.?) and the Type II error rate p_1 (i.e. ?) will be functions of n.

Alternately, ? and ? can be specified and the sample size n determined accordingly. We will assume this design procedure is being used and develop expressions for n and c.

In this context, a Type I error means that an unimpaired stream is declared to be impaired. Thus

$$\begin{aligned} ? \; ? \; & P(\bar{y} ? c | p ? p_0) \\ ? \; & P(\bar{y} ? c | ? ? ?_0) \\ ? \; & P ? \frac{\bar{y} ? ?_0}{\frac{?}{\sqrt{n}}} ? \frac{c ? ?_0}{\frac{?}{\sqrt{n}}} ? \\ ? \; & 1 ? P ? Z ? \frac{c ? ?_0}{\frac{?}{\sqrt{n}}} ? \end{aligned}$$

and

$$c ? ?_0 ? \frac{?}{\sqrt{n}} z_{1 ? ?}$$

A Type II error occurs when an impaired stream is declared to be unimpaired. When $p = p_1$,

$$? \; ? \; P(\bar{y} ? c \,|\, p ? \; p_1)$$
$$? \; P(\bar{y} ? c \,|\, ? \; ? ?_1)$$
$$? \; P\left(\frac{\bar{y} ? ?_1}{\frac{?}{\sqrt{n}}} ? \frac{c ? ?_1}{\frac{?}{\sqrt{n}}} \right)$$

$$? \; P\left(Z ? \frac{c ? ?_1}{\frac{?}{\sqrt{n}}} \right)$$

and

$$c ? ?_1 ? \frac{?}{\sqrt{n}} z_? \; .$$

The required sample size can be found by solving for n in the above equations. This gives

$$n ? \frac{(z_? \; ? \; z_{1??})^2}{(z_{1? p_0} \; ? \; z_{1? p_1})^2} \, .$$

Example

Suppose a particular sample will be declared unacceptable if the level of a certain pollutant is larger than $U = 1000$ *ppm*. Suppose $? = 100$ *ppm*. Let $p_1 = 0.10$ be the acceptable proportion and let $p_1 = 0.25$ be the unacceptable proportion. Find the sample size required to ensure $? = ? =$ 0.10.

$?_0 \; ? \; 1000 ? 100 z_{0.9} \; ? \; 1000 ? 100 ? 1.282 \; ? \; 871.8\,ppm$ and $?_1 \; ? \; 1000 ? 100 z_{0.75} \; ? \; 1000 ? 100 ? 0.675 \; ? \; 932.5\,ppm$

The required sample size is $n ? \dfrac{(z_{0.10} ? z_{0.9})^2}{(z_{0.9} \; ? \; z_{0.75})^2} ? 18$ and the corresponding value of c is

$c ? 932.5 ? \dfrac{100}{\sqrt{18}} ? z_{0.1} \; ? 902.3$. The site should be declared impaired if the sample mean obtained from eighteen observations is greater than 902.33 *ppm*.

It may not be feasible to obtain a sample $n = 18$ observations. That number can be reduced by increasing $?$ or $?$. If we let $? = ? = 0.25$ then only $n = 5$ observations are required.

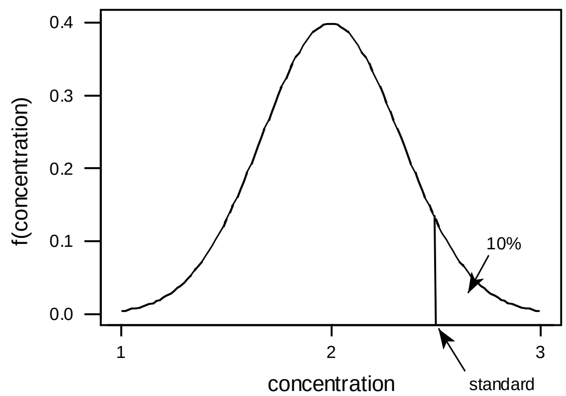

Figure 1. Plot of distribution of hypothetical chemical concentration. The standard allows for exceeding a concentration of 2.5 10% of the time.

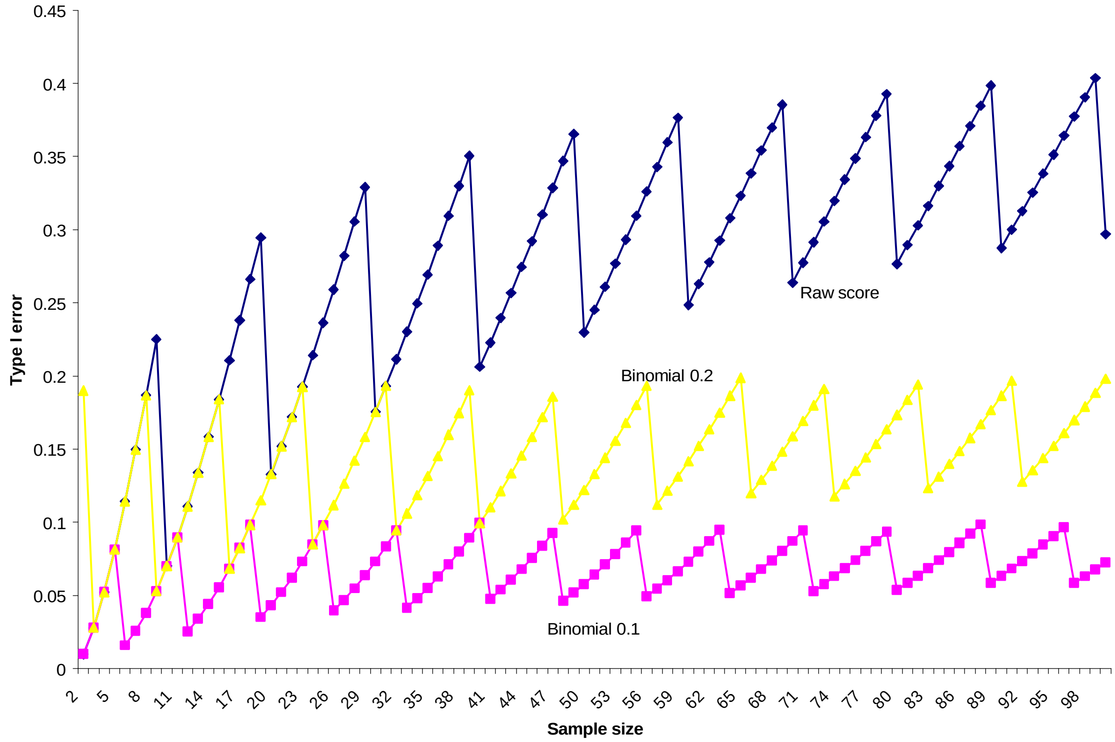

Figure 2. Type I error rates (false declaration of impairment) for Binomial and Raw score methods for varying sample sizes. The Type I error rate for the Binomial is set at 0.1 or 0.2.

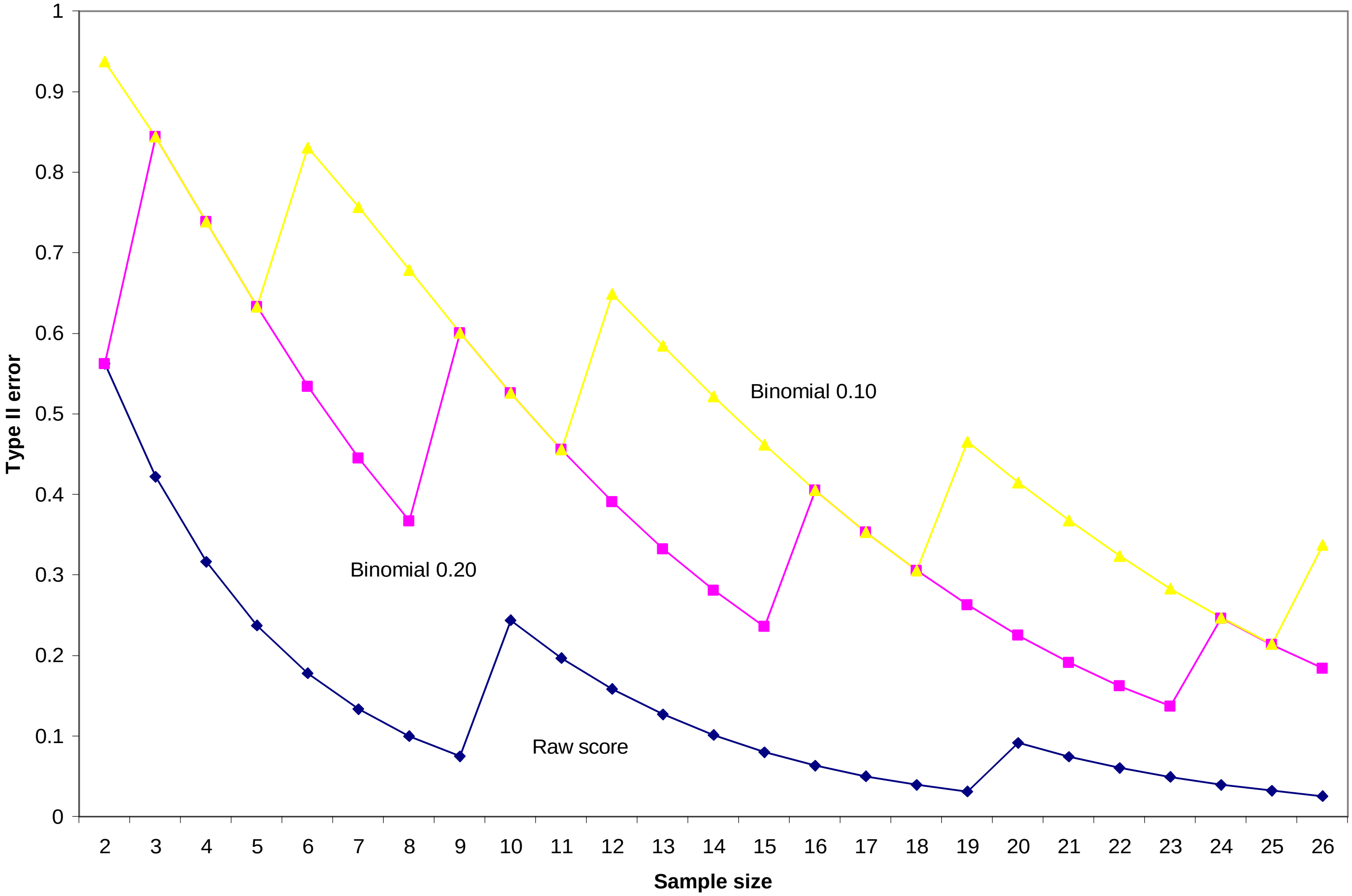

Figure 3. Type II error rates for Binomial and Raw score methods for sample sizes from 2 to 25. Binomial method uses at Type I error rate of 0.10 or 0.20

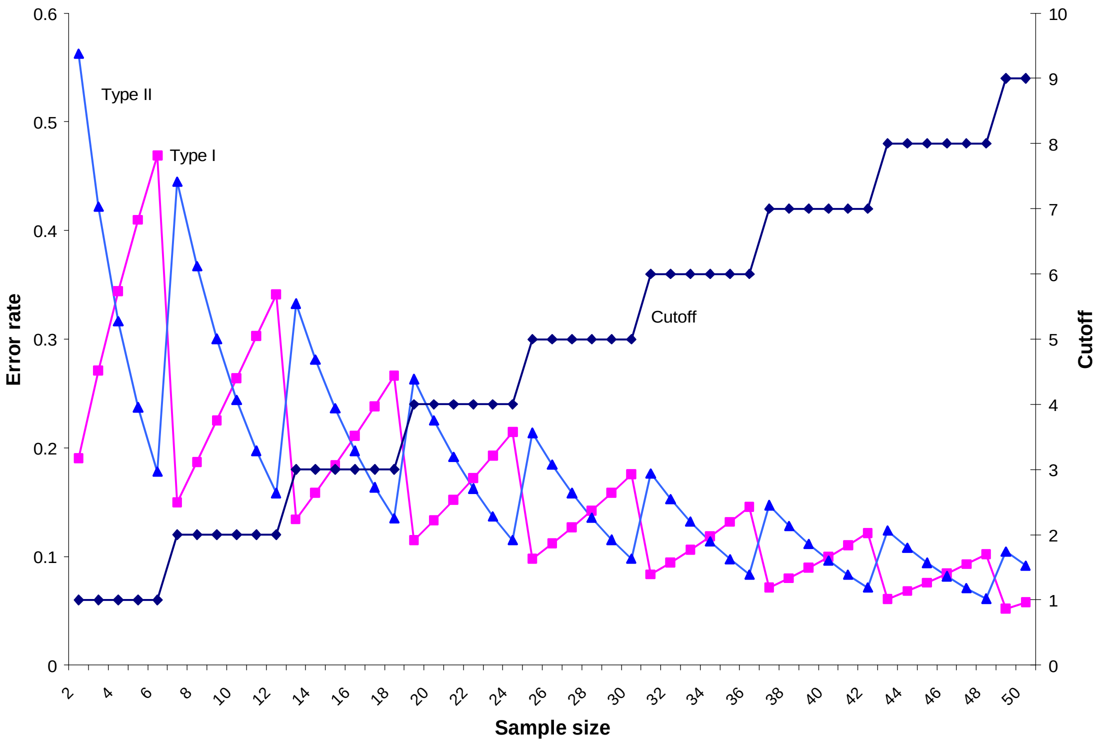

Figure 4. Error rates and cutoffs for different sample sizes, trying to make the Type I and Type II error rates as close as possible for the Binomial test. Cutoff values correspond to the minimum number of samples that may exceed the standard to declare the site as impaired.

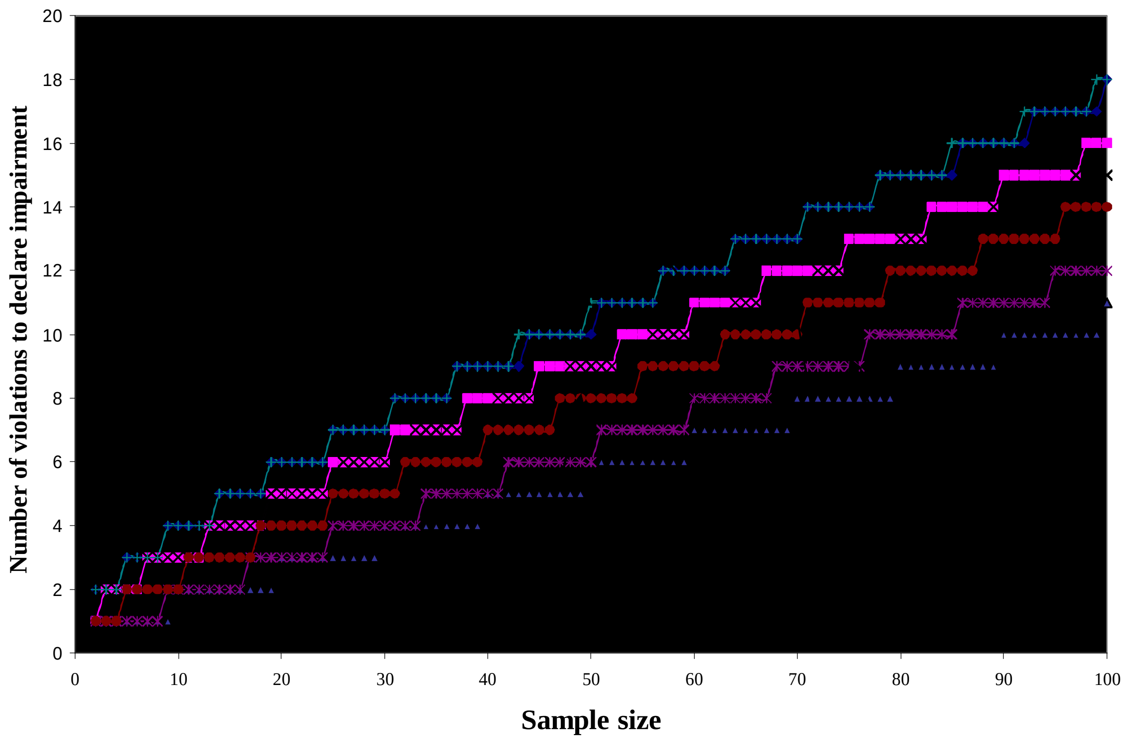

Figure 5. Number of violations required to declare a site as impaired for different sample sizes.

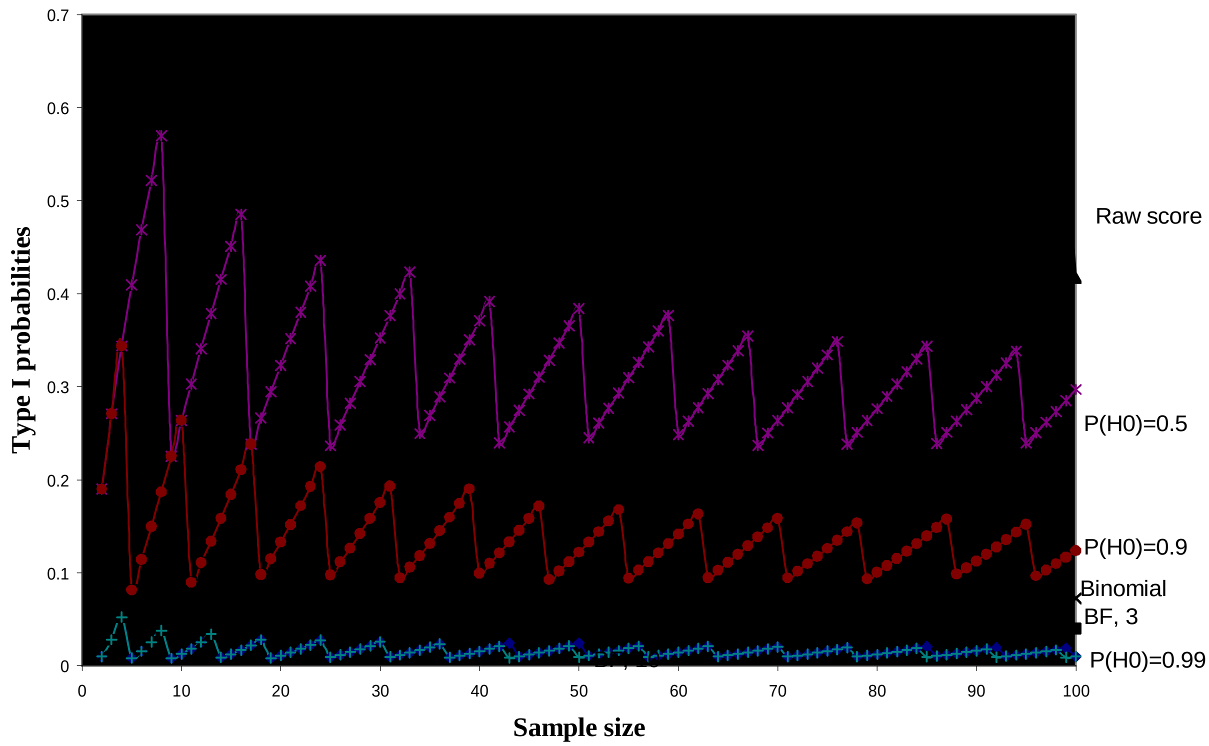

Figure 6. Type I probabilities for various methods. Binomial method is based on setting the Type I error rate at 0.1.

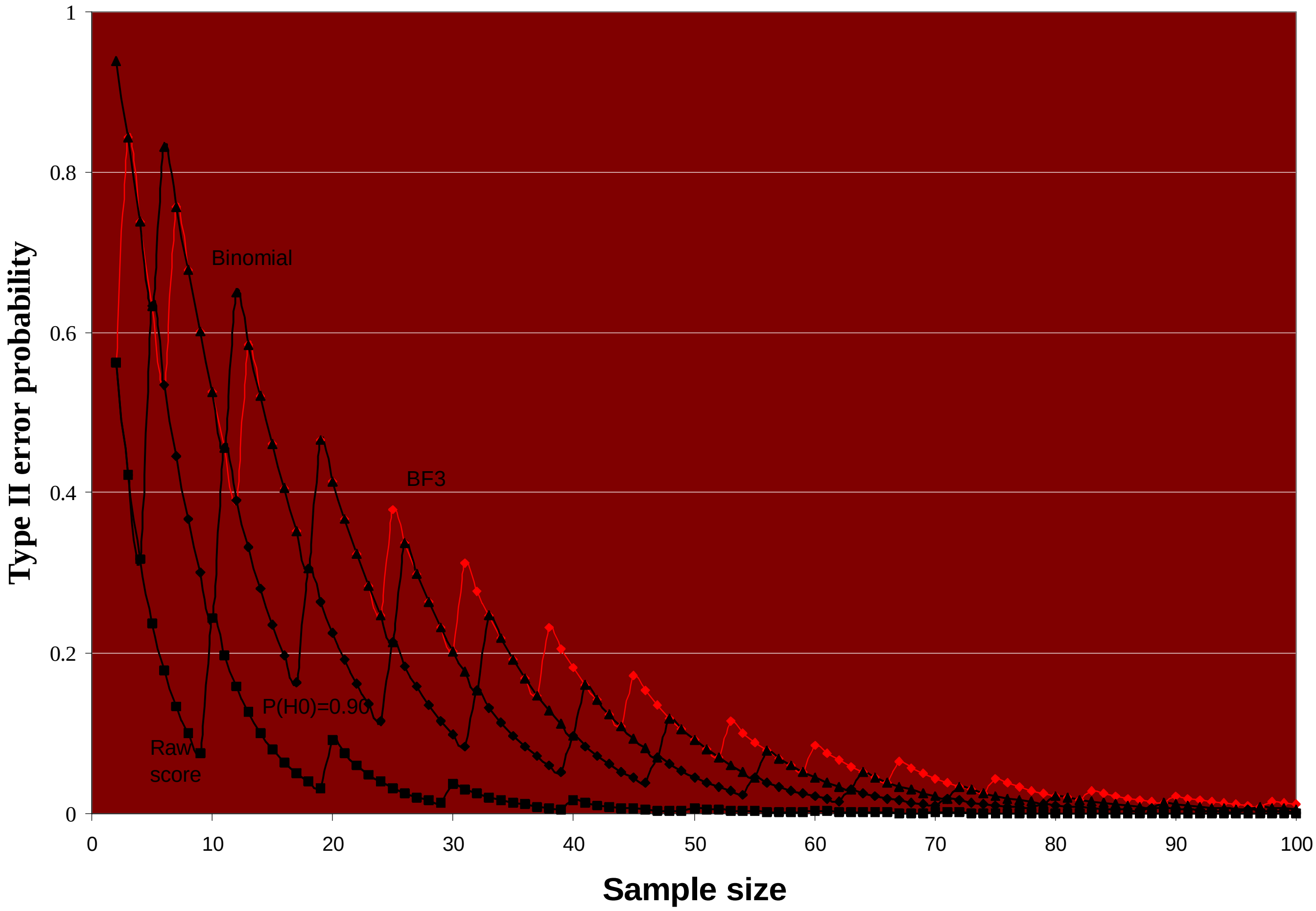

Figure 7. Type II probabilities for various methods. P(H0)=0.9 represents the Bayesian method with a prior of the null hypothesis set at 0.9, BF3 refers to the Bayes Factor method using 3 as a cutoff.

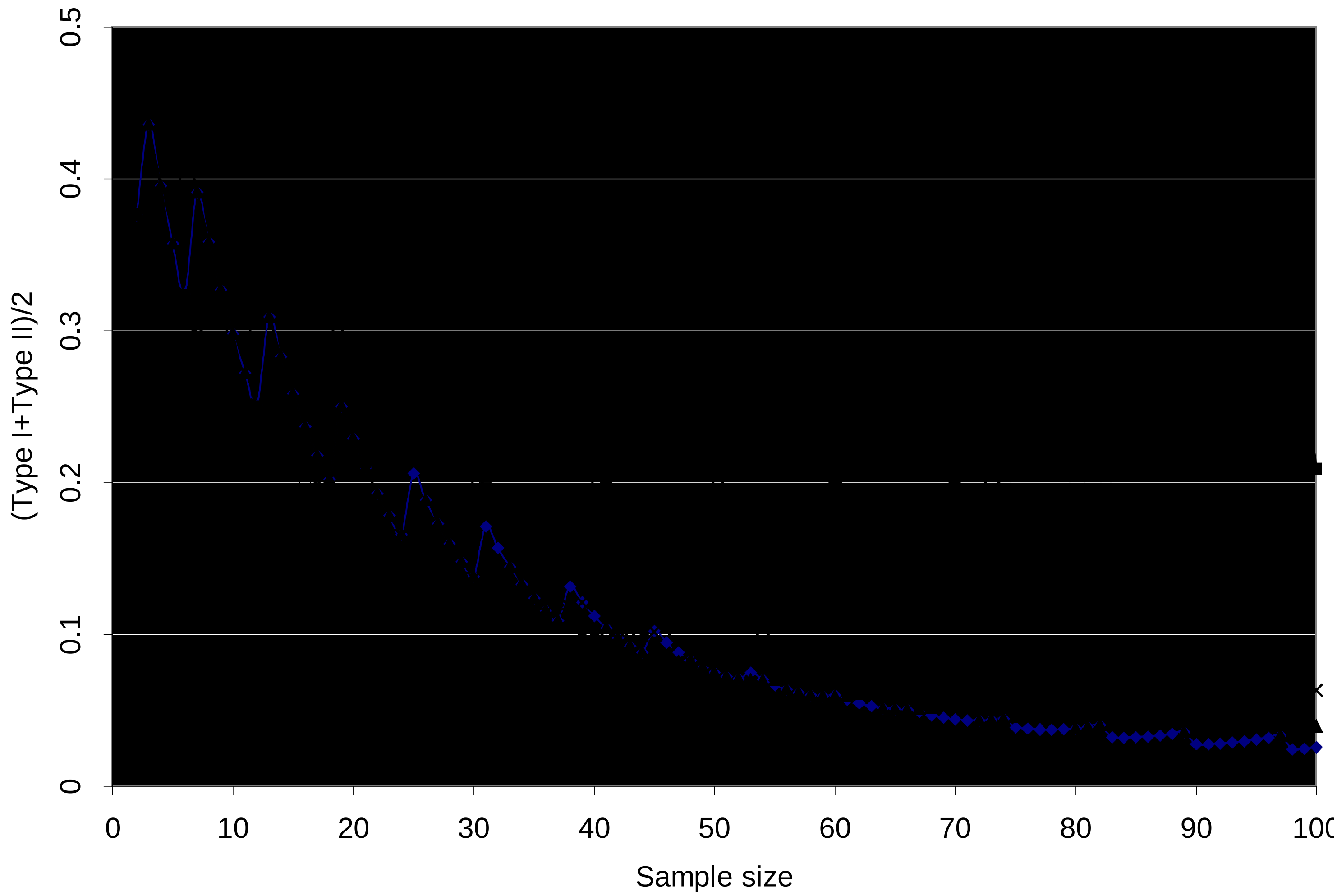

Figure 8. Average error rate of the different methods using different sample sizes. P(H0)=0.9 represents the Bayesian method with a prior of the null hypothesis set at 0.9, BF3 refers to the Bayes Factor method using 3 as a cutoff.

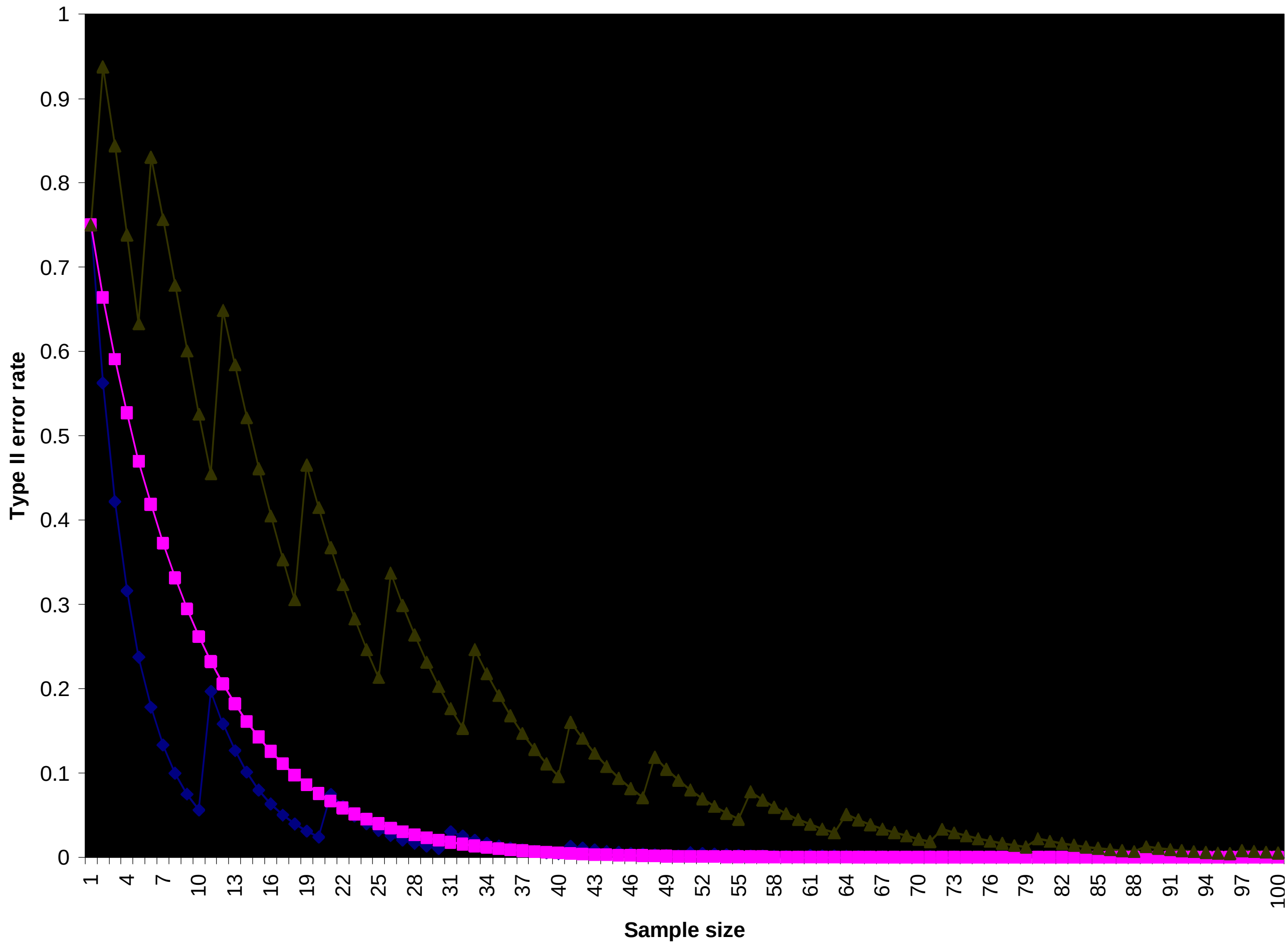

Figure 9. Plot of Type II error rates for Acceptance sampling, compared with Raw score and Binomial methods.

4. Virginia's Water Quality Standards: Future Development Issues and Recommendations

Prepared by Carl Hershner

Introduction

Section 303 (c) of the Clean Water Act requires that State and Tribal standards protect public health or welfare, enhance the quality of water and serve the purposes of the Act. As defined in Sections 101 (a), 101 (a)(2) and 303 (c) of the Act, "serve the purposes of the Act" means that water quality standards should: 1) include provisions for restoring and maintaining chemical, physical and biological integrity of State waters, 2) provide, wherever attainable, water quality for the protection and propagation of fish, shellfish and wildlife and recreation in and on the water ("fishable/swimmable"), and 3) consider the use and value of state waters for public water supplies, propagation of fish and wildlife, recreation, agricultural and industrial uses and navigation. A water quality standard consists of designated uses that describe the existing and/or potential uses of a water body, water quality criteria to protect the designated uses, and an antidegradation policy to maintain and protect both existing water quality and existing uses.

Use Designations

Use designation in the water quality standards programs of the states of Maryland, North Carolina and the Commonwealth of Pennsylvania were reviewed and compared to Virginia's program. As expected, the general structure of statements water quality standards, including Maryland, North Carolina, Pennsylvania and Virginia is similar. However, states outside Virginia appear to have more efficiency and flexibility in application of their water quality standards than Virginia.

The Virginia standards include a limited number of designated uses which have been assigned to all waters combined with an out-of-date classification system. Virginia has 7 classes of waters, plus special designations for public water supply, shellfish, nutrient sensitive waters, and trout waters. Virginia include classes that are simply geographic location (Classes I-IV, ocean, estuarine, non-tidal and mountain) and classes which can be considered uses (Class V, stockable trout and Class VI, natural trout). Virginia has designated all waters with the uses of recreation and propagation of aquatic life, wildlife and edible and marketable natural resources (9VAC25-260-10). However, under the current program, water quality criteria are assigned to class groups, saltwater or freshwater, (which actually splits Class II) with additional criteria for special designations.

Pennsylvania and Maryland have designated uses for waters and have assigned appropriate criteria to protect the designated use(s). Somewhat of a hybrid, North Carolina's program varies in that criteria are assigned to classes that have designated uses. North Carolina ranks waters according to class (Classes WS, B and C) for purposes of assigning general water quality criteria and subclasses within the water supply class allow for the designation of land use control regulations criteria. Pennsylvania has the most sophisticated program of those reviewed. Pennsylvania's water quality standards program has 15 uses. While most of the criteria, particularly for toxic substances, are the same for all uses, dissolved oxygen and temperature criteria vary according the sensitivity of the designated use. All three states, to varying degrees, have assigned different criteria to protect different designated uses.

There are several advantages to restructuring Virginia's use designations. Elimination of the classification system and specification of designated uses with appropriate criteria would simplify the regulations, making them less cumbersome and easier to interpret for both the regulated community and the program managers. By mirroring the federal program structure, Virginia would be in a better position to respond to existing federal guidance and requirements and address new federal initiatives (nutrient criteria, biocriteria, and TMDL's) resulting from the Clean Water Action Plan. Perhaps most importantly, the simplified structure would allow for the flexibility to assign criteria according to sensitivity of use(s), protecting sensitive uses with the most stringent criteria, and less sensitive uses with other criteria.

Converting Virginia's program to a more straightforward system of hierarchical standards based on specific use designation would not be difficult. A first step might be review of the uses suggested by the federal program, to select those appropriate to Virginia's waters and management objectives. This selection would necessarily have to consider any existing use designations. In addition, existing regulatory or legislative guidance on other uses or values attributed to Virginia's waters should be considered. As previously noted, Virginia has already assigned all Commonwealth waters "fishable/ swimmable" uses. The special designations of public water supply, shellfish waters, and trout waters (Classes V and VI) are also, effectively, uses. Guidance on other potential use designations can be found in the State Policy as to Waters, which includes the following language:

> "Beneficial use" means both instream and offstream uses. Instream beneficial uses include, but are not limited to, the protection of fish and wildlife habitat, maintenance of waste assimilation, recreation, navigation, and cultural and aesthetic values. Offstream beneficial uses include, but are not limited to, domestic (including public water supply), agricultural, electric power generation, commercial and industrial uses. Public water supply uses for human consumption shall be considered the highest priority' (Va Code § 62.1-10).

If implemented, this cumulative list could provide the flexibility to assign a basic set of standards to protect the basic "fishable/swimmable" uses, and a hierarchy of increasingly restrictive standards to protect other more sensitive uses. This does not necessarily represent a wholesale redevelopment of Virginia's water quality standards. Rather, it would be a reorganization of the existing program to eliminate class designations. In place of these designations, individual water bodies or reaches would be directly assigned uses. The state would then be in a position to consider the application of use-attainability analysis in the TMDL process or in other water quality management settings. In the end, specific water quality parameter criteria could then be directly linked to specific uses and plans made to assure that those uses are attained. The benefit is primarily clarity in program application.

Developing a clear linkage between use designation and water quality standards should facilitate evaluation of attainable use designation and/or the consequences of any new use designation. Clear indication of water quality conditions which must be maintained/attained to support each potential use, should make the management program much more comprehensible to managers (particularly local government officials) and the regulated community. At this point the state would be simply making the link between standards, uses, and individual water bodies clearer.

The direct utilization of use designation then could facilitate a continuing program of water quality standard review. The criteria necessary to support each use could be periodically evaluated with reference to the evolving science, pollution control technology and public preferences. The rationale for any change in water quality standards would then be logically and intuitively linked to both the best technical information and the intended uses of any affected water body.

Water Quality Criteria

Toxic pollutants

As required under Section 303 (c)(2)(B) of the Clean Water Act, states are required to adopt criteria for all section 307(a) priority toxic pollutants for which the Environmental Protection Agency has published criteria. If data indicate that one of the 126 toxic pollutants identified by the EPA will interfere with the attainment of a designated use, then the state must adopt a numeric criterion for the specific pollutant. As such, the EPA, under the auspices of the Clean Water Act, is the lead authority for toxicant criteria. The federal effort relieves Virginia of the responsibility for development of specific toxic pollutant criteria.

This is not to suggest Virginia should never review the criteria. The Commonwealth needs to evaluate which toxic pollutants are relevant to each of the designated uses of state waters. This should be an ongoing process, which seeks to incorporate the best available technical information. The determination of specific numeric criteria, however, is not left to the state. There is significant federal technical effort committed to evaluation and development of criteria which the state can utilize. As a result in depth technical review of numeric toxic criteria would not seem to be a pressing an issue, as some of the other existing and potential standards.

Other Pollutants

Criteria for the other, standard pollutants in water quality management (e.g. temperature, pH, DO, BOD, suspended solids, fecal coliforms, etc.) are set at levels which are generally accepted. Virginia may benefit from a specific technical review of these criteria, however, if it restructures its program to more clearly implement specific use designation. It would be appropriate to implement a program of routine periodic review of the standards with a specific focus on the criteria established for each designated use. The goal should be to ensure that the numeric criteria are consistent with the best available technical information.

There are some aspects of the current standards which, in the current litigious climate of environmental management, may warrant fresh consideration. The issue is often not the water quality objective, but rather the method of specifying acceptable compliance. Mixing zone criteria seem to be a particular example. In the specific case of thermal pollution, while the standards are sufficient to guide hydrodynamic modeling of discharge limits, there is only a poor relationship to the actual behavior of real systems. Additionally, monitoring programs are in no way related to the specified performance criteria. The point is that a review of some of the criteria might lead to more appropriate and beneficial guidance. Considering potential rather than theoretical impacts, and accepting the realities of monitoring programs, might lead to suggestions for practical alternatives to current criteria.

Fecal Coliform

It is noteworthy that there continues to be a significant debate about the application of fecal coliform standards. This issue is sufficiently substantive, and has enough potential significance for TMDL implementation, to warrant an immediate and focused review of the state of the art and the options for water quality standard development. Even if the state does not have the option to make immediate changes in these criteria, it should engage in an assessment of the issues and the implications for Virginia's water quality management program. This will provide the necessary background for any policy review/debate which may become appropriate.

Biological Criteria

Perhaps the biggest need for technical input to Virginia's water quality standards program arises in the development of biological criteria. EPA is pressing this issue. Indeed, EPA is actively engaged in a wide variety of research efforts to develop technical bases for biological criteria which states might implement. Virginia needs to be proactive in this issue, not just to be prepared to respond to EPA. Biological criteria may be the only practical solution to the need for water quality criteria in a significant proportion of the Commonwealth's waters. In particular, the ocean, estuarine, and wetland classes of waters currently designated by Virginia, all have natural conditions which defy efforts to develop criteria under the standard list of water quality parameters. The most practical solution for these classes (and/or for any uses designated in these types of waters) may be development of biological criteria.

Much of the work on biological criteria for water quality have been undertaken in free flowing, freshwater streams. While these efforts have produced a variety of suggested parameters which integrate physical, chemical, and biological conditions with some consistency - the future of biological criteria is much broader. EPA is sponsoring research on the development of indicators for the ecological integrity of wetlands (one of the "waters" of the Commonwealth). Pennsylvania has continuing research on biological indicators for assessing degree of disturbance in wetlands, streams, and watersheds. The Chesapeake Bay Program is pursuing identification of "biological endpoints" to establish water quality restoration goals in the estuarine waters of Virginia and Maryland.

The pressure to undertake development of biological criteria is evident in EPA's designation of the Chesapeake Bay and major tributaries in Virginia as impaired waters. This will ultimately force the state to develop management plans, including TMDLs, to improve water quality.

Determination of what needs to be accomplished will inevitably be based on a living resource rationale - and that will be a de facto biological criterion.

The point is that Virginia has every reason to actively pursue development and implementation of biological criteria. The Commonwealth will not be able to avoid biological criteria in management of its estuarine waters thanks to EPA. It will also probably discover, much like Pennsylvania, that biological criteria are the only effective way to manage wetland waters. Indeed, biological criteria may well be the practical solution to management of other "problem" waters such as lakes, seeps, mine drainage, bogs, and "flashy" watercourses. The state of the science is such that useful guidance can be developed for a number of these types of waters. For others, a commitment to pursue current consensus and invite periodic review of biological criteria, will be the best practical alternative to inappropriate application of numeric criteria for standard pollutant parameters.

Recommendations

Three activities are relevant to Virginia's water quality standards program:

1. Restructure the program to eliminate "class" designations of water, replacing that with a straightforward "use" designation. (A workshop to recommend specific criteria for each of the existing and potential uses of Virginia waters would be a first step.)

2. Develop biological criteria for as many of Virginia's waters (or "uses" if the first recommendation is implemented), as practical. (Some preliminary synthesis of the state of the science, followed by a series of water/use specific workshops, might produce some useful guidance.)

3. Establish a formal program for the periodic review of all the standards and a protocol of use-attainability analysis to introduce site specific conditions that may alter the standard to be achieved. (Preparation of review/discussion papers focused on particular standards or groups of standards, followed by a workshop might be one strategy for providing technical guidance and consensus.)

Appendices

Appendix A. Comparison of Uses by State
A table comparing the uses of state waters designated or otherwise implemented in Virginia, Pennsylvania, North Carolina, and Maryland.

Appendix B. North Carolina Water Quality Standards

Appendix C. Maryland Water Quality Standards

Appendix D. Pennsylvania Water Quality Standards

Appendix A. Comparison of Uses by State

	STATE			
USES	Virginia	Pennsylvania	North Carolina	Maryland
Public Water Supply	Public Water Supply	Public Water Supply	Classes WS I-V	P (additional designation)
Fish and Wildlife	Fish and Wildlife	Warm Water Fish	Class C Aquatic Life Secondary Rec.	Aquatic Life and Water Contact
	Threatened and Endangered	Aquatic Wildlife Supply		
		Migratory Fish		
Native Trout	Natural Trout	Cold Water Fish	Tr (trout)	Natural Trout
Stocked Trout	Trout Stocked Fishery	Trout Stocked Fishery	Tr (trout)	Recreational Trout
High Quality	High Quality	High Quality	High Quality	
Exceptional Value	Exceptional Value	Exceptional Value	Outstanding Resource	
Recreation	Recreation	Boating	Class C Boating	
		Fishing	Class C Fishing	
		Water Contact	Class B Water Contact	
Commercial	Commercial/ Industrial	Industrial		
Shellfish	Shellfish			Shellfish
Cultural /Aesthetic	Cultural /Aesthetic	Aesthetic		
Agriculture	Agriculture	Livestock		
		Irrigation		
Navigation	Navigation			
Electric Generation	Electric Generation			
Waste Assimilation	Waste Assimilation			
Nutrient Sensitive			Nutrient Sensitive	

Appendix B. North Carolina Water Quality Standards

CLASS	STANDARD								
	Chlor A	DO	Solids	Coliform	pH	Temp	Turbidity	Chemical	Special
Freshwater C	<40 ug/l	<5 mg/l daily		200/100 ml or 20 % 400/100 ml	6-9	not to exceed 5 above normal	<50 NTU stream <25 NTU lakes	see table below	
WS I			<500 mg/l	50/100 ml					natural watershed public ownership
WS II								same as WS I	undeveloped stormwater, buffer and veg.
WS III-V									
Saltwater C	<40 ug/l	<5 mg/l			6.8-8.5	No increase >0.8 summer 2.2 winter	<25 NTU		
SA shellfish No sewage discharge				Median not to exceed 14/100 10% exceedance 43/100					
SB primary contact				same as C					
Nutrient Sensitive									
High Quality									Discharge limits BOD_5=5 mg/l NH_3-N=2 mg/l DO =6 mg/l volume<50% 7Q10
Outstanding Resource									Fish habitat or fisheries recreation special designation state park/forest RTEs, regulates marinas

North Carolina Water Quality Standards (continued)

	Fresh Class C	WS I-V	Trout Waters	Saltwater C
Arsenic	50 ug/l			same
Beryllium	6.5 ug /l	6.8 ng/l		Ns
Cadmium	0.2ug/l		0.4 ug/l	5.0 ug/l
Chromium	50 ug/l			20 ug/l
Cyanide	5 ug/l			1 ug/l
Flouride	1.8 mg/l			NS
Lead	25 ug/l			25 ug/l
Methylene	0.5 mg/l			NS
Mercury	0.012 ug/l			0.025ug/l
Nickel	88 ug/l	25 ug/l		8.3 ug/l
PCB	0.001 ug/l	same	same	same
Selenium	5 ug/l			71 ug/l
Toluene	11 ug/l		0.36 ug/l	NS
TBT	0.008 ug/l			0.002ug/l
Aldrin	0.002 ug/l	0.127 ng/l		0.003 ug/l
Chlordane	0.004 ug/l	0.575 ng/l		same
DDT	0.001 ug/l	0.588 ng/l		same
Demeton	0.1 ug/l			same
Dieldrin	0.002 ug/l	0.135 ng/l		same
Endosulfan	0.05 ug/l			0.009 ug/l
Endrin	0.002 ug/l			same
Guthion	0.01 ug/l			same
Heptachlor	0.004 ug/l	0.208 ng/l		same
Lindane	0.01 ug/l			0.004ug/l
Methoxychlor	0.03 ug/l			same
Mirex	0.001 ug/l			same
Parathion	0.013 ug/l			0.178 ug/l

Toxaphene	0.0002 ug/l			same
Copper	7 ug/l			3 ug/l
Iron	1.0 mg/l			Ns
Silver	0.06 ug/l			0.1 ug/l
Zinc	50 ug/l			86 ug/l
Chloride	230 mg/l	250 mg/l		
Barium		1.0 mg/l		
Manganese		200 ug/l		
Nitrate		10 mg/l		
2,4-D		100 ug/l		
2,4,5-TP		10 ug/l		
Sulfates		250 ug/l		
Benzene		1.19 ug/l		
C tetrachloride		0.254 ug/l		
Cl Benzenes		488 ug/l		
Dioxin		0.000013 ng/l		
Hexachloro butadine		0.445 ug/l		
PAH		2.8 ng/l		
Tetrachloroethane		0.172 ug/l		
Tetrachloro ethylene		3.08 ug/l		
Vinyl chloride		2 ug/l		

Minimum Standards are assigned to Freshwater Class C and Saltwater Class C. More stringent standards are assigned by Class in order to protect the designated use (s)

Appendix C. Maryland Water Quality Standards

USE	CRITERIA					
	DO	Coli.	pH	Temp	Turbidity	Toxic Substance
I contact aquatic life	<5 mg/l daily	200/100 ml or 10 % 400/100 ml	6.5-8.5	not to exceed 90	<150 NTU maximum <50 monthly mean	criteria to protect aquatic organisms, fish consumption
I-P contact aquatic life water supply	same as I	50/100 ml	same as I	same as I	same as I	aquatic organisms, fish consumption, water supply
II shellfish	same as I	<14 MPN /100ml	same as I	same as I	same as I	aquatic organisms, fish consumption
III Natural Trout Chlorine discharge prohibited	<5 mg/l minimum, <6 mg/l daily mean	same as I	same as I	< 68 or ambient	same as I	aquatic, organisms, fish consumption
III-P Natural Trout, Water Supply	<5 mg/l minimum, <6 mg/l daily mean	same as I	same as I	< 68 or ambient	same as I	aquatic, organisms, fish consumption, water supply
IV Rec. Trout	same as I	same as I	same as I	<75 or ambient	same as I	aquatic organisms, fish consumption
IV-P Rec. Trout, Water Supply	same as I	same as I	same as I	<75 or ambient	same as I	aquatic organisms, fish consumption, water supply

Appendix D. Pennsylvania Water Quality Standards

CRITERIA FOR ALL USES

Aluminum		>0.1of 96-hour LC_{50}
Alkalinity $CaCO_3$		< 20 mg/l
Ammonia Nitrogen	Formula based on pH & temp	
Bacteria		Summer: < 200/100ml, 2,000/100 ml
Fluoride		F_1 2 mg/l and F_2 0.01 of 96-hr LC_{50}
Iron		1.5 mg/l total, <0.3 mg/l dissolved
Mn		<1.0 mg/l
N		<10 mg/l nitrogen
Osmotic Pressure	<50 milliosmoles /kg	
pH		6-9
Phenolics		$Phen_1$ <0.005 mg/l, $Phen_3$ mean 0.02 mg/l
Solids		500 mg/l mean 750 mg/l max
Total Residual Chloride		4 day 0.011 mg/l, 1 hr. 0.019
DO		>6.0 daily
	5.0 min	>5.0 daily
	4.0 min	>6.0 February-July
Temp		monthly criteria <87

CWF- Cold Water Fishes
WWF- Warm Water Fishes
MF- Migratory Fishes
TSF- Trout Stocking
PWS-Public Water Supply
IWS- Industrial Water Supply
LWS- Livestock water Supply
AWS-Wildlife Water Supply
IRS Irrigation
B-Boating
F-Fishing
WC- Water Contact
E-Esthetics
HQ- High Quality
EV- Exceptional Value
N– Navigation

5. Review of the Commonwealth's Biological Monitoring Program

Leonard Smock

The Biomonitoring Subcommittee of the Water Quality Academic Advisory Committee (WQAAC) was formed to provide assistance to DEQ in their efforts to evaluate and update the Commonwealth's ongoing biological monitoring program. Five faculty members from Virginia academic institutions were appointed to the committee based on their ongoing involvement with biomonitoring issues. A list of the members is attached.

The subcommittee met with DEQ personnel on May 18, 1999 at DEQ headquarters in Richmond to initiate its work. A list of meeting attendees is attached. At that meeting, DEQ personnel provided draft working documents that outlined and provided insight into many of the issues that DEQ perceives are central to the further development of its biomonitoring program. DEQ personnel briefly presented each issue and all members participated in ensuing discussions that facilitated a better understanding of the issues and objectives.

The subcommittee has identified a number of issues, listed below, that it presently is addressing. Each subcommittee member was designated as the lead individual for developing a review of several issues. The subcommittee will submit a draft report, projected for summer 2000, to the full WQACC for review prior to its submission to DEQ. DEQ can then use that report to assist it in developing its biomonitoring strategies and work plans.

Points identified for review:

1. Extent of the ongoing biomonitoring program.
   Is there a need for DEQ to expand or contract its present biomonitoring effort?

2. Issues of adopting a Probabilistic Biomonitoring program.
   What would be the general structure of such a program? Benefits? Costs? How would it complement the existing program?

3. Methodology to be used and the need for program modernization
   Address the need for updating of biomonitoring methods to conform to latest EPA protocols. Focus on sampling procedures, level of taxonomy required, an ecoregion approach, the choice of metrics and their calibration, and data analysis techniques. Develop an action plan to adopt the required methodology.

4. What are the appropriate organisms to monitor?
   Is there a need to add another group of organisms, such as fish, to the biomonitoring program? What are the costs and benefits?

5. QA/QC issues

Issues concerning sampling protocols, accuracy of taxonomy, the data base, etc. Is there a need for development of a training program for DEQ personnel? Is there a need for and feasibility of a certification program for DEQ personnel?

6. Resource issues
   What resources issues need to be addressed for DEQ to appropriately accomplish the present and recommended biomonitoring effort? This includes resources needed for DEQ to comply with the Clean Water Act requirements for determining the biological condition of all streams in the Commonwealth.

7. Need for and role of a toxicity testing program
   Is there a need for DEQ to develop a toxicity testing program (or other program) to complement the benthic invertebrate monitoring program that would allow DEQ to determine the cause of impairment that is detected by the monitoring program? Should such a program be limited geographically (e.g. to the Bay region) or be implemented statewide?

8. Need for a wetlands biomonitoring program
   Is there a need for DEQ to develop a wetlands biomonitoring program?

9. Extent and need of an estuarine monitoring program
   Is there a need to further expand the existing estuarine monitoring program and integrate with the freshwater biomonitoring program?

10. What is the potential role or citizen biomonitoring efforts in relation to DEQ biomonitoring?

An additional issue, concerning the evaluation of present biocriteria methods used by DEQ, initially was to be addressed by this Biomonitoring Subcommittee. Instead, it now will be addressed in a broader sense by the entire WQACC.

Members of the Biomonitoring Subcommittee

Dr. Greg C. Garman
Department of Biology and Center for Environmental Studies
Virginia Commonwealth University
Richmond, VA 23284
Ph. 804-828-7202
Fax 804-828-0503
ggarman@atlas.vcu.edu

Dr. R. Chris Jones
Department of Biology
George Mason University
Fairfax, VA 22030
Ph. 703-993-1127
Fax 703 –993-1046
rcjones@gmu.edu

Dr. Eric P. Smith
Department of Statistics
VPI&SU Tech
410 Hutcheson Hall (0439)
Blacksburg, VA 24061
Ph. 540-231-7929
Fax: 540-231-3863
E-mail: epsmith@vt.edu

Dr. Leonard A. Smock
Department of Biology
Virginia Commonwealth University
Richmond, VA 23284
Ph. 804-828-1562
Fax 804-828-0503
lsmock@saturn.vcu.edu

Dr. J. Reese Voshell
Department of Entomology
VPI&SU
Blacksburg, VA 24061
Ph. 540-231-5707
Fax 540-231-9131
rvoshell@vt.edu

Attendees of the May 18, 1999 meeting
of
DEQ personnel and Biomonitoring Subcommittee members

Greg Garman	VCU	804-828-7202
ggarman@atlas.vcu.edu		
Wick Harlan	DEQ – TRO	757-518-2154
wtharlan@deq.state.va.us		
Chris Jones	GMU	703-993-1127
rcjones@gmu.edu		
Mark Richards	DEQ – CBP.	804-698-4392
marichards@deq.state.va.us		
Lou Seivard	DEQ – WQS	804-698-4121
ldseivard@deq.state.va.us		
Tony Silvia	DEQ – TRO	757-518-2140
agsilvia@deq.state.va.us		
Don Smith	DEQ – Monitoring Coordin.	804-698-4429
dhsmith@deq.state.va.us		
Eric Smith	VPI&SU	540-231-7929
epsmith&vt.edu		
Len Smock	VCU	804-828-1562
lsmock@saturn.vcu.edu		
Reese Voshell	VPI&SU	540-231-5707
rvoshell@vt.edu		
Larry Willis	DEQ – WCRO	540-567-6825
ldwillis@deq.state.va.us		

6. Bio-monitoring – The New Water Quality Management Challenge

Leonard A. Smock

Presentation made at the Virginia Water Resources Research Symposium, November 15, 1999, Richmond, VA

The past decade has seen an increased use of bioassessments as part of routine water quality monitoring programs. This increase is driven by regulatory statues that suggest or require the incorporation of biological data in monitoring programs and the recognition by environmental managers of the added value that biological information provides to them concerning a broad spectrum of water quality issues. Bio-assessments include a broad range of methods such as bioassays, analysis of contaminant uptake by target species, and biological monitoring efforts.

Biomonitoring can be defined as the continuing collection of data to establish if specific water quality conditions and designated uses of waterways are being met. Biomonitoring relies on the premise that aquatic species or communities integrate the short- and long-term effects of a wide variety of perturbations. In doing so, biomonitoring provides direct information on the biological health of a waterway, which often is a key issue for water quality assessments. Biomonitoring typically is used in a variety of ways, including as an early warning system of changes in water quality, to detect episodic perturbations, for targeted site assessment studies, and for area-wide trend assessments.

In Virginia, trend assessment focuses on streams and rivers and includes the efforts for statewide monitoring for 305(b) and 303(d) reporting. The Virginia Department of Environmental Quality routinely monitors about 180 sites semiannually. Benthic macro-invertebrates are the organisms of choice for the state's monitoring effort, which among other objectives, are used to directly assess the ability of streams and rivers to meet the "aquatic life" designated use. The data from the monitoring program is analyzed and distilled into the following narrative ratings of water quality for the 305(b) and 303(d) reports: fully supporting, fully supporting but threatened, partially supporting, and not supporting.

A variety of issues have arisen as biomonitoring has matured into a fully accepted and often required water quality assessment tool. These issues are being discussed among state and federal water quality agencies as well as by the scientific community. The issues include those that have arisen as science has provided more information on the validity and methodology for biomonitoring, regulatory agencies have issued guidelines for biomonitoring protocols, and environmental policy has increasingly incorporated the routine use of biomonitoring by water quality managers.

An issue that has recently arisen and that presently faces Virginia regulatory agencies is the choice of methods to be used in statewide biomonitoring efforts. The Environmental Protection Agency (EPA) published biomonitoring protocols in 1989 that were widely adopted by many states. Those protocols have undergone extensive use, testing, and analysis since that time and those efforts quickly showed the need for changing the protocols. The EPA has recently modified the protocols and biomonitoring programs should be moving to adopt the new procedures detailed in the document. Among the most important changes incorporated in the new protocols is that of requiring the metrics to be used in the data analysis to be chosen from a suite of potential metrics rather than having the protocols stipulate that a given set of metrics must be used. Preliminary studies will be necessary to both determine which metrics are most appropriate to detect changes in water quality in Virginia's waterways and also to then calibrate those metrics with the appropriate narrative water quality ratings. Determining the thresholds that separate metric values into the four narrative ratings for 305(b) and 303(d) reporting is a difficult but critical aspect of establishing a scientifically and legally defensible biomonitoring program.

Another important aspect of any biomonitoring effort, and one that is critical for statewide biomonitoring, is that of establishing and adhering to a set of standard methods. A number of issues concerning standard techniques often arise when implementing a biomonitoring effort and the EPA protocols provide considerable latitude in the specific methods to be used. Included among these are questions concerning which organisms are to be monitored, sampling gear, which habitats are sampled, field sampling methods, the level of taxonomy used in identifying organisms, and methods of data analysis and presentation.

One question initially facing biomonitoring programs is "which organisms do we monitor?" As noted, many programs use benthic macroinvertebrates as the organism of choice for a variety of reasons. The question of the level of taxonomic resolution to be used for these organisms presently is a matter of debate that includes both scientific and resource issues and which does not seem to have a clear resolution. A well-informed decision on the level of taxonomy to be used needs to be made at the initiation of a biomonitoring program and standardized throughout the entire program.

A stringent QA/QC program focused on these and other pertinent aspects of the program are just as important as the question of appropriate metrics and threshold values in making the biomonitoring program scientifically and legally defensible. Training and certification programs may need to be implemented to provide appropriate QA/QC throughout the program.

Organisms, other than macroinvertebrates, also can be employed in biomonitoring programs. Standard methods exist for fish biomonitoring efforts, and the importance of the ability of the public to identify with fish more so than with invertebrates should not be underestimated. Since the EPA now suggests that more than one type of organism should be used in biomonitoring programs, program managers need to consider their need and ability to incorporate both macroinvertebrates and fish into routine biomonitoring efforts.

A problem facing biomonitoring programs, and particularly those using benthic macro-invertebrates, is that of finding suitable reference sites. Biomonitoring with macroinvertebrates typically requires that aspects of the macro-invertebrate communities at monitored sites be compared with those at a reference site. Ideally, the reference site would have a level of water, habitat, and biological characteristics that we would like to have in the monitored sites. But what constitutes a reference site? To some, a reference site should reflect “pristine” conditions, but there are few streams remaining in Virginia that could be held to this high standard. Most areas of the state do not have any streams that even come close to being in a pristine condition. More commonly, we use streams that are considered the “best available,” but which often have characteristics that are far from those that would be found in a pristine stream.

Given the problem of a lack of pristine streams, different states use different criteria to select reference streams. Many establish *a priori* water, habitat, riparian zone, and watershed conditions that a stream must meet before it is considered as a reference site. Other states, including Virginia, rely on “best professional judgement,” whereas some are moving to incorporate the use of model reference streams based on a probabilistic analysis of data derived from a number of high-quality sites.

No matter how the sites are chosen, agencies and individuals overseeing biomonitoring efforts must recognize how critical the quality of reference sites is to the final product of biomonitoring, i.e., that of determining the quality of the monitored waterway. The higher the quality of the reference site, the more difficult it is for the monitored site to attain a high narrative water quality rating. On the other hand, if a reference site with relatively low quality is used (which too often is the condition of “best available” sites), attaining a higher narrative rating for a monitored site is much easier. The selection of reference sites thus becomes an issue that is based on both science and policy. This issue must be clearly understood because it has important environmental and economic policy consequences.

A final point that is pertinent to both the problem of converting to new EPA biomonitoring protocols as well as to the need for appropriate reference sites is that of the geographic variability in the characteristics of waterways across the state. The concept of ecoregions is widely used in biological monitoring to account for geographic variability among waters of different geographic areas. The waters of different ecoregions typically have different physico-chemical, hydrological, and biological characteristics that can cause considerable differences among ecoregions in both the appropriate metrics to be used and in the expected reference site metric scores.

Virginia encompasses six ecoregions, but with sufficient variability in some of them to require the recognition of subecoregions to adequately separate groups of similar streams. If the problem of the need for adequate reference sites is to be addressed, it will need to be done within the framework of ecoregions and subecoregions. In addition, the selection of appropriate metrics and the calibration of those metrics, as required for the new EPA protocols for benthic macroinvertebrates, will need to be accomplished within an ecoregion framework. Metrics that are appropriate in one ecoregion may not be so in another and a given value of a metric in one ecoregion may not correspond to the same narrative water quality rating in another. An integrated analysis to determine the appropriate metrics, and to calibrate those metrics, across all ecoregions in Virginia thus is required before the new EPA protocols can be fully adopted.

Biomonitoring is a useful tool for the assessment of water and habitat quality in streams and rivers and is especially applicable for determining if the “aquatic life” designated use is being met. A key consideration for state environmental agencies and others establishing new or continuing ongoing biomonitoring programs is to implement standardized programs that are in step with accepted protocols. There are a number of other issues that need to be considered and resolved early in the process, some of which have been discussed above. If these issues are recognized and dealt with, biological monitoring programs will become an important tool for environmental managers and regulators in the quest for preserving and improving the quality of our streams and rivers.

7. Addressing the Issue of Fecal Indicators in Virginia's Waters

Prepared by Howard Kator

We have identified a number of issues and problems related to Virginia regulatory compliance, monitoring and research programs. To address these issues, there is a need for in-depth discussion among scientists and agency experts. To achieve this objective, we recommend that a workshop consisting of scientists with expertise in indicator microbiology and public health water quality microbiology be convened. Participants will be expected to submit position papers on selected contemporary topics in environmental and sanitary microbiology. These papers will be edited and published as a VWRRC peer-reviewed publication. This publication would also contain recommendations addressing what Virginia must do to insure its sanitary microbiological programs continue to provide the most meaningful and effective protection of the public health and natural resources. The workshop will be co-sponsored by VWRRC/DEQ/VIMS.

The focus of the workshop could be on all or perhaps two of the following current topics:

1. Are currently approved microbial indicators valid measures of sanitary water quality?

Questions to be posed include: Do these indicators predict the presence of fecally-derived human pathogens in receiving waters? There are two currently approved indicators' systems which are based on fecal coliforms and the enterococci. Is there a role for these indicators in regulation of receiving waters? Do numerical standards based on these indicators accurately reflect health risk? And if the validity of these indicators is questionable what alternatives exist? Is the enterococci based indicator system of EPA superior to the fecal coliform? What needs to be done in the future?

2. Is currently approved methodology effective?

A significant link exists between indicators and available methodology for reasons that include: effects of environmental conditions on their recovery from natural waters; physiological status, culturable vs. nonculturable, and continuing need for inexpensive techniques that yield rapid results. Are current approved methods for recovery of viable cells effective? If not, how can they be improved and what alternative methods appear most promising?

(A literature review is needed to summarize improvements in existing approved methods, to identify new methods, both culturable and molecular, and new sampling techniques.)

3. What methods are now available for identification of sources of fecal indicators? Do these methods require validation? What are the most important questions that future research should ask? What is required to apply these approaches in Virginia?

Associated with the EPA concept of the Total Maximum Daily Load (TMDL), is the notion of remediation through intervention and source identification. Source identifciation requires differentiation of human from animal contamination and in many areas, of animal from animal. Several methods based on ribotyping, antibiotic resistance profiles, or use of indicators with inherent source-specificities have been proposed. These should be identified, discussed and future research needs and likelihood of success considered.

4. Can methods for the direct detection of pathogens provide a valid means to assess water quality and public health risk? And if so what are the most promising approaches?

Various investigators have proposed that direct detection of pathogens replace surrogate methods based on indicators. Questions arise concerning the public health significance of such data and how such data can be interpreted. Indicators have been used for reasons that include difficulty of pathogen detection, transient nature of pathogens, and lack of theoretically establishing predictive relationship for all pathogens and disease risk. Is the indicator concept still a viable one? How can direct pathogen detection be incorporated into a regulatory framework?

Proposed workshop participants:

Keynote- DEQ Director?

Alpha Diallo, Norfolk State Laboratory, City of Norfolk
George Simmons, Virginia Tech
Charles Hagedorn, Virginia Tech
Howard Kator, VIMS
Shaban Kotob, VIMS
Martha Rhodes, VIMS
Bruce Wiggins, James Madison University
Cameron Hackney, Virginia Tech
VDH representative
DEQ representative
DCR representative
State Epidemiologist
NVPDC representative (local government)

Tentative Schedule:

1. Form a steering committee of three to identify participants, focus on topic areas, evaluate proposed products, cost, etc.
2. Pick date for workshop, location, establish work schedule.
2. Hold workshop, conduct literature review, prepare and edit proceedings and papers.
4. Submit for publication by the VWRRC

8. Sampling and Monitoring Design for 305(d) Program – Combining Probabilistic and Judgmental Sampling

Eric P. Smith

The Virginia Department of Environmental Quality (DEQ) must address questions of trend in major pollutants as well as the extent of pollution. The current monitoring program for rivers and streams is designed to address questions about change in levels of measurements over time (e.g. trend) or to compare upstream and downstream sites (e.g. change). The selection of sites is based on what is referred to as judgment sampling. In judgment sampling, sites are selected based on the opinion of experts. Sites are typically selected based on location of industrial sites, sewage treatment plant location and reference locations. This type of sampling may be quite good for trend evaluation at particular sites and for assessment of improvement in water quality at the site. However, the design may not give adequate estimates of the extent of pollution. The estimates are appropriate for the sites selected and not the complete set of sites that represent the state's aquatic resources. The common approach to obtaining unbiased estimates of quantities such as the proportion of impaired sites is based on probabilistic sampling.

The EPA has suggested that probability based designs be used to estimate important environmental quantities. Given that current programs are based on judgment sampling, an important question is whether or not the current program can be combined with a probabilistic sampling program. Below some approaches to combining the data sets and assumptions required are discussed.

Two main approaches have been suggested to combine information from different sampling programs. Overton et al. (1993) describe an approach based on matching. A second approach is suggested by Rathbun (1999) and is based on spatial-temporal correlation among observations from probabilistic and judgmental samples.

The Overton method is based on characteristics of the possible samples representing the site. The collection of all possible samples is referred to as the sampling frame. First, the sampling frame is created. Each site has certain characteristics such as the location, depth, habitat factors, land use factors, etc. The second step is to cluster or group the set of possible sites (from the frame) based on these attributes. Then the judgment samples are assigned to the clusters. The judgment sample is then assumed to be representing that cluster and would be treated as a probabilistic sample. To be effective, the judgment samples must be representative. If not, the estimates made from using these samples will be biased and adjustments are required.

The approach suggested by Rathbum (1999) is to use the judgment sample to estimate data that would have been observed from the start of the probabilistic program. Data is collected at the same time or same time interval from both the probabilistic sample and the judgment sample. These samples are then related through their correlation. The spatial and temporal relationship between the samples is used to develop a model that allows prediction of environmental quantities at new sites. Assuming the correlation model is valid, the judgment samples at previous times may then be used to estimate samples that would have been collected from a probabilistic sample at that time. The method requires that the spatial model does not change over time. A bias correction for bias due to judgmental sampling is also required.

Of the two approaches, the method based on matching would seam more appropriate to the needs of the DEQ and easies to implement. An approach that may merit more attention is to view the fixed sampling sites as representing a subpopulation of the state's resources. The probabilistic sample is then chosen to augment this subpopulation and to make the overall sample representative (i.e. it fills in the gaps). The interesting problems are how to choose the probabilistic sample given the information in the fixed sample, how large a sample is required and how to combine the information together. An approach to the combination of samples is through weighing. A sample is weighted according to the information that it represents in the population.

References Cited

Overton, J., T.C. Young, and W.S. Overton. 1993. Using 'Found' Data to Augment a Probability Sample: Procedures and Case Study. Environmental Monitoring and Assessment 26: 65-83.

Rathbum, S.L. 1999. Sampling Design Issues for Section 305(b) Water Quality Monitoring. Unpublished material.

9. Modeling and the Alternatives for Water Quality Management

Theo Dillaha and Shaw Yu

During the past winter and spring, Drs. Dillaha and Yu reviewed topics related to TMDL development and implementation and then met on May 7, 1999 in Charlottesville to compare notes and identify specific topics to be addressed. On March 23, 2000 a larger group of the Academic Advisory Committee met via a phone conference and additional topics were discussed. Many of the TMDL modeling topics identified overlap with priority issues of other Water Quality Academic Advisory Subcommittees.

Topics to be addressed and some preliminary recommendations:

1. Existing techniques and models for estimating pollutant loading to receiving waters do not consider the effects of spatial variability in upland conditions well. This makes it difficult to evaluate the benefits of buffer zones, ponds, detention structures, and other structures whose performance is site specific. The ability for consider spatial variability in upland areas is also critical for identifying critical source areas so that pollution control resources can be targeted specifically to the areas causing water quality problems.

2. The existing approach used to simulate bacterial transport in the BASINS (HSPF) model is weak. HSPF simulates bacteria as a dissolved pollutant, which does not interact with sediment. Research in Virginia and elsewhere indicates that bacteria are partially adsorbed and that deposited sediment in streams is a reservoir for bacteria. Release of fecal coliform bacteria from stream sediments during low flow periods could be the source of frequent stream standard violations during low flow periods. To address this, bacteria should be simulated in both the free and adsorbed phases. HSPF has this capability, but research is needed to develop appropriate model parameters.

3. Available water quality models simulate chemical and physical water quality parameters plus bacteria and chlorophyll-A, a surrogate for phytoplankton. To develop TMDLs for benthic impairment, correlations must be developed between the parameters the models simulate and resulting benthic impairments. This will enable us to evaluate alternative upland management practices that may be required to control benthic impairments. (Also addressed by Subcommittee: Bio Monitoring and Assessment.) We are on the program committee of a conference on Benthic Impairment and TMDLs that will be held in Charlottesville on June 7. We will be meeting with scientists and agency personnel involved in bethic assessment and restoration at this meeting and will attempt to refine our goals and objectives at this meeting and to find additional collaborators.

4. Virginia water quality standards appear to be unrealistic for some streams. Some streams may need to be reclassified for uses other than contact recreation (also addressed by Priority Issue 2 Subcommittee: Water Quality Standards for Water

Quality Assessment). The recently completed fecal coliform TMDL studies for Mill Creek and Pleasant Run in Rockingham County, found that wildlife contributions alone were sufficient to violate the 30-day geometric mean fecal coliform standard.

5. One hundred percent attainment is not realistic for nonpoint sources that may be dominated by extreme events like hurricanes. A similar approach to that used with point sources (e. g., 7-day, 10-year low flow) may be required. For example, water quality standards could be based on 30-day geometric mean during extreme flow periods or instantaneous standards would only have to be met 99% of the time. (Also addressed by Water Quality Standards for Water Quality Assessment.)

6. Monitoring programs, which collect monthly samples, do not provide good estimates of 30-day geometric mean concentrations that are required for development of TMDLs for fecal coliform bacteria. Monitoring protocols need to be revised (new quality assurance project plans) to insure that the data collected supplies the information needed to determine compliance with 30-day geometric mean standards. If the state monitoring program cannot determine compliance with the standard, the standard should be abandoned. (Opportunities for Improved Application of Statistical Procedures on 303(d) and 305(b) Reports.)

7. Costs of TMDL development and implementation in Virginia could be lowered if TMDL development tasks common to all TMDL studies were consolidated. For example, development of weather data files is essential for TMDL development. Existing weather station data in Virginia is incomplete. Consequently, each TMDL study contractor must spend considerable time and resources assembling and integrating the required weather records from multiple weather stations to obtain a complete weather record for the study area. This is a complex task that could be more efficiently accomplished if one contractor was given the task of creating complete weather records for all weather stations in Virginia. Similar savings might also be achieved through the development of a comprehensive inventory and database of available water quality and flow records.

8. If there were nutrient standards many streams might have multiple impairments. If nutrient standards are issued TMDL studies for different impairments on the same stream should be coordinated and addressed simultaneously rather than letting separate contracts for each individual impairment. This will greatly reduce TMDL development costs.

9. Every fecal coliform TMDL study conducted in Virginia has suggested that the primary source of water quality impairment is cattle in streams and all have concluded that the removal of cattle from the stream is required to eliminate the impairment. The state, with EPA concurrence, should consider extrapolating these results to all fecal coliform impaired streams rather than spending millions of dollars on additional studies that reach the same conclusions.

10. Currently, separate contracts for TMDLs for streams with multiple impairments are being prepared for each impairment. This is inflating TMDL development costs. Part of the problem is that there are many streams with both fecal coliform and

benthic impairment. Until the benthic standards and a benthic TMDL procedures are secured no TMDLs studies be conducted on streams with benthic impairments. When a benthic TMDL procedure is developed, TMDLs for streams with benthic and other impairments, should be conducted under one contract. This should significantly reduce costs assuming that the contractor is technically capable of doing TMDLs for benthic and other impairments.

Appendix A: *Proposed* Fiscal year 2001 Work Plan of the Water Quality Academic Advisory Committee

The FY 2001 work is organized into three areas.

Area 1: Presentations to EPA Region III

Area 2: Continuation of working group on bio-assessment.

Area 3: For benthic and fecal impairments, review of monitoring, standard setting and assessment protocols and recommend actions to better coordinate these three areas of agency activity.

Area 1: EPA, Region 3 and Headquarters, has expressed interest in a presentation on some of the statistical work that has been done under the AAC umbrella and other work on trends that DEQ has contracted through the Center. A presentation to EPA would involve Leonard Shabman, Eric Smith Carl Zipper and Golde Holtzman. We have already had a conference call with region 3 on the bi-nominal method. The trend analysis project work can be logically made part of the same presentation.

Deliverable Products:

1. A one day presentation workshop with DEQ and EPA staff in Philadelphia on the statistical work will be completed as soon as schedules permit.

2. A report revised and adapted from the 1999-2000 report on the binomial method for review by the EPA headquarters, if requested by region 3. This report will be finalized within 30 days after the meeting with EPA; a draft report will be prepared in advance of the meeting.

Area 2: The Biomonitoring Subcommittee of the Water Quality Academic Advisory Committee (WQAAC) was formed to provide assistance to the DEQ in their efforts to evaluate and update the Commonwealth's ongoing biological monitoring program. Five faculty members from Virginia academic institutions were appointed to the subcommittee based on their ongoing involvement with biomonitoring issues. In 1999 the subcommittee identified a number of issues that it would address. Each subcommittee member was to be designated as the lead individual for developing a review of several issues. The subcommittee would submit a draft report to the full WQACC for review prior to its submission to the DEQ. The DEQ would then use that report to assist it in developing its biomonitoring strategies and work plans. This work will be completed under this FY 2001 work plan. (note: Some of the work under area 2 will contribute to the effort under area 3 below).

Deliverable Product:

3. A report from the AAC to DEQ on biomonoitoring will be prepared. The report will address the following topics: 1) The need for the DEQ to expand or contract its present biomonitoring effort. 2) The general structure of a probabilistic biomonitoring program and how it would complement the existing program. 3) Biomonitoring techniques used and the need to conform to the latest EPA protocols. 4) Determination of appropriate organisms to monitor. 5) The need for development of a training program for the DEQ personnel on sampling protocols, accuracy of taxonomy, the database, etc. 5) Resources needed for DEQ to comply with the Clean Water Act requirements for determining the biological condition of all streams in the Commonwealth. 6) Need for, and geographic location of, toxicity testing to complement the benthic invertebrate monitoring to assist in determining the cause of benthic impairment. 7) Need for a wetlands biomonitoring program. 8) Need to expand the existing estuarine monitoring program and integrate it with the freshwater biomonitoring program.9) Critique the existing narrative biocriteria adopted by the state and required by EPA. 10) Role of citizen biomonitoring efforts in relation to the DEQ biomonitoring?

Area 3: At the present time the DEQ is organized with separate agency units responsible for water quality standards, monitoring and assessment for 305(b) and 303 (d) reporting. There is a logical link among the three units in that monitoring is undertaken and assessment is the process of establishing whether water quality standards are met. Historically, however, monitoring and standard setting were only loosely linked by the assessment process. Monitoring resources were often limited and sample sizes were small. What was monitored and where the monitoring occurred was governed by criteria other than how the monitoring would serve a statewide watershed assessment process, as now called for. Standards have been established, but these often are narrative (qualitative), may not be linked to whether monitoring data would (or could) be collected and may not be sensitive to the unique circumstances that might exist in different areas of the state. The most obvious example of the failure to relate monitoring, standards and assessment (now being remedied) is that dissolved oxygen standards did not recognize that some swamp water areas would never achieve the DO concentrations called for by the standards. Monitoring went ahead and the assessment reported DO levels that did not meet the state standard that had been developed for faster moving streams and rivers in other geographic areas. The assessment process had to report that the swamp waters did not meet the DO standard and then explain why the state did not list the waters as impaired (i.e. the DO standard should not apply). However, there was no alternative standard for these waters so there was no benchmark that assessors could use for the considering the monitoring data.

A less obvious, but perhaps more significant, problem is that the cost of assessing waters as impaired is becoming quite significant. The emerging importance of the 303(d) listing process and the prospect that impaired waters (waters that do not meet standards) will need to be subjected to expensive TMDL planning and implementation, has focused increased attention on the scientific basis for the standards, the nature of monitoring data and the proper interpretation of those data and standards. The setting of standards, the protocols for monitoring and the decision criteria for making the assessment need to recognize the limited resources available for water quality improvement programs. Limited resources will be best allocated when inappropriate or unattainable standards are identified, when monitoring programs are as accurate as possible and when assessment and planning studies use scientifically defensible tools for data evaluation.

At present most waters in the Commonwealth are listed as impaired for violating either the existing fecal water quality standard or a benthic condition. Observers of the assessment and 303 (d) listing process have commented to DEQ that, for benthic impairments, the state lacks proper criteria for making an impaired waters listing. They note that the narrative criteria that are used are inadequate for a regulatory process. Meanwhile others have noted that the criteria for listing stream segments for fecal contamination should be reviewed in light of questions about the use of the fecal coliform count as an impairment indicator when the sources are not known and the consequences for ecological and human health are not certain.

Therefore, the AAC will, for benthic and fecal issues, review and critically evaluate monitoring, water quality goals (achievable and designated uses and standards) and assessment protocols (including techniques for relating stressors to outcomes) and recommend actions to better coordinate these three areas of agency activity. The goal of the review and evaluation is to help the DEQ to make the most cost effective use of limited monitoring, planning and implementation funds to secure water quality improvements in the state.

The DEQ will provide relevant DEQ documents relating to fecal contamination and benthic impairments to the AAC for its review, including but not limited to guidelines for the assessment process, guidelines for monitoring of fecal and benthic conditions, statements of the relevant standards that have been employed in the assessment process and any internal agency or external reviews of any of these topics.

The Center will initiate the review by developing a draft conceptual statement of the problem of integrating monitoring, standards and assessment protocols for the DEQ with the objective of making the most cost effective use of limited monitoring, planning and implementation funds to secure water quality improvements in the state. This draft concept will be discussed and modified at a one day meeting of the AAC and DEQ staff. Using the conceptual statement as a guide the AAC will review relevant documents and report to the DEQ areas where improvements can be readily made and areas where more long term research of policy development may be necessary.

Deliverable Product:

4. A report will be submitted to the DEQ by the AAC that will, for benthic and fecal issues, review and critically evaluate monitoring, water quality goals (achievable and designated uses and standards) and assessment protocols (including techniques for relating stressors to outcomes) and recommend actions to better coordinate these three areas of agency activity.