COMMENTS SUMMARY REPORT

External Peer Review of
EPA Analysis of Epidemiological Data From EPA Bacteriological Studies

Contract No. 68-C-02-091
Versar Work Assignment No. 1-11

Prepared for:

U.S. Environmental Protection Agency
Office of Water
Office of Science and Technology
Health and Ecological Criteria Division
301 Constitution Ave, N.W.
Washington, D.C. 20004

Prepared by:

Versar, Inc.
6850 Versar Center
Springfield, Virginia 22151

February 2004

TABLE OF CONTENTS

I. INTRODUCTION

The United States Environmental Protection Agency (EPA), Office of Water is charged with protecting public health and the environment from adverse exposure to chemicals and microbials in water media, such as ambient and drinking waters, waste water/sewage sludge and sediments. In support of this mission, the Office of Water/Office of Science and Technology (OST) develops health standards, health criteria, health advisories, and technical guidance documents for water and water-related media.

In 1986, EPA published *Ambient Water Quality Criteria for Bacteria*–1986. That document contained EPA's recommended water quality criteria for bacteria to protect bathers from gastrointestinal illness in recreational waters. The water quality criteria identified levels of indicator bacteria, namely *Escherichia coli* (*E. coli*) and enterococci, that demonstrate the presence of fecal pollution and which should not be exceeded to protect bathers in fresh and marine recreational waters. Indicator organisms such as these have long been used to protect bathers from illnesses that may be contracted from recreational activities in surface waters contaminated by fecal pollution. These organisms generally do not cause illness directly, but have demonstrated characteristics that make them good indicators of harmful pathogens in waterbodies. Prior to its 1986 recommendations, EPA recommended the use of fecal coliforms as an indicator organism to protect bathers from gastrointestinal illness in recreational waters. However, EPA conducted epidemiological studies and evaluated the use of several organisms as indicators, including fecal coliforms, *E. coli*, and enterococci, and subsequently recommended in 1986 the use of *E. coli* for fresh recreational waters and enterococci for fresh and marine recreational waters because they were better predictors of acute gastrointestinal illness than fecal coliforms. Some states and authorized tribes have replaced their fecal coliform criteria with water quality criteria for *E. coli* and/or enterococci; however, many other states and authorized tribes have not yet made this transition.

In the 1986 criteria document, EPA recommended the use of a risk level associated with 8 illnesses per 1000 swimmers in fresh waters and 19 illnesses per 1000 in marine waters. This represents approximately a 1-2% risk that recreators will suffer from gastrointestinal illness from swimming in ambient recreational waters. These risk levels were identified based on the concentrations of *E. coli* and enterococci that roughly correlated to the previous fecal coliform criterion. However, EPA believes that it is appropriate for states and authorized tribes to exercise their risk management discretion when protecting recreational waters. Based on a review of the studies used in the derivation of EPA's §304(a) criteria for bacteria, EPA recommends states and authorized tribes select a risk level for fresh waters between 0.8 and 1.0 percent. However, some have suggested that EPA may be recommending risk level bounds for freshwater that are too restrictive given the type of data and analysis performed, particularly given the risk level bounds recommended for marine waters.

Under this work assignment, Sections 1.5 and 1.6 of the "*Implementation Guidance for Ambient Water Quality Criteria for Bacteria*" were externally reviewed by a panel of three peer reviewers. The peer reviewers were asked to address major issues associated with the approach used to determine the appropriate risk level range for recreators in fresh waters. The peer review panel included experts in statistical analysis, particularly those associated with microbiology and

epidemiological studies. The three reviewers were Joseph Eisenberg, Charles McGee, and Mark Sobsey.

II. CHARGE TO THE PEER REVIEWERS

The peer reviewers were charged with responding to the following Technical Charge:

1. Given the constraints of the data available, is the risk analysis in the *Implementation Guidance for Ambient Water Quality Criteria for Bacteria* appropriate?

2. Is it scientifically defensible to extrapolate the relationship (in terms of linear regression or other quantitative means) between bacterial indicator density and illness rate for fresh waters beyond the 1% risk level?

3. How much further could one extrapolate and what would be the rationale for extrapolating further?

III. GENERAL COMMENTS

Joseph Eisenberg

I have some questions on the use of geometric means for estimating the dose of exposure. It is the arithmetic mean that provides the appropriate average exposure over time? The geometric mean, which is a better estimate of the median, will tend to underestimate the average level of exposure.

It would be nice to see Figures 1.3 and 1.4 for enterococci in fresh and marine waters.

IV. RESPONSE TO CHARGE

1. Given the constraints of the data available, is the risk analysis in the "Implementation Guidance for Ambient Water Quality Criteria for Bacteria" appropriate?

Joseph Eisenberg
The guidelines are appropriate as written precisely because they do not go beyond the limits of the data. See answers to 2 and 3 for further clarification.

Charles McGee
For many years, this analysis has been the subject of much debate. However, the experimental design, the quality of the data gathered and duplication of the results by other researchers has made the risk analysis put forth in the guidance defensible.

A current challenge to the original research upon which the risk analysis is based is whether the spatial and temporal variability of the beach water quality was captured in the experimental design. In any study, the strength of the relationships between two variables is dependent on the precision of the measurement of those variables. EPA's own EMPACT study and research on recreational water contamination carried out on the west coast has demonstrated the significance of this variability. In preparing my answer to this question, I reviewed some of the original EPA publications, and I was convinced that the study design adequately addressed this concern. A second issue that should be addressed is how measurement error was taken into account in the calculation of risks criteria for indicator bacteria densities (see Table 4 in EPA440/5-84-002 [*Ambient Water Quality Criteria for Bacteria - 1986*]). Specifically, the question was whether water quality variation can be better explained by systematic patterns or is variability simply the result of random error. Dr. Joon Ha Kim and I had considered this point in a journal article evaluating error associated with California's marine water quality monitoring and public notification procedures. Upon submission of this article to the journal, one of the article's reviewer's comments triggered a rewrite of the article. The concepts in that rewrite not only apply to the expression of error in the formula in Table 4 but to all three questions posed in this peer review. Because of time constraints, I could not contribute to the rewrite and had my name removed as an author. I have spoken with Dr. Kim, and he has agreed to allow you to contact him for an advance copy of his new article. He would also like to suggest some other data analyses of the EPA and Santa Monica Bay data. Dr. Kim's address is:

Joon Ha Kim, Ph.D.
944A Engineering Tower
Chemical Engineering & Materials Science
University of California, Irvine
Irvine, California 92697-2575
Phone number 949-824-7754

[Note: articles provided by Dr. Kim are provided in Appendix B, following Charles McGee's comments.]

Mark Sobsey

The answer is "no". The reasons for this answer will be given below [see comments under "Specific Comments" section] in more detail.

2. ***Is it scientifically defensible to extrapolate the relationship (in terms of linear regression or other quantitative means) between bacterial indicator density and illness rate for fresh waters beyond the 1% risk level?***

Joseph Eisenberg

No. The linear regression model is a data driven model; i.e., there is no mechanistic rationale for the model structure. The predictions beyond the data, therefore, are not reliable or defensible.

One potential way to extrapolate is to assume the sigmoidal dose-response relationship, based on data from dosing trials. These trials have been conducted for various pathogens and administer higher doses than observed in epidemiology studies. Given these trial data, one can assume that the relationship is linear at low doses, increases exponentially at higher doses, and eventually saturates at yet higher doses. This is basically what is said in the document under review, and is illustrated in Figure 1.2. The problem becomes where to place the cut point. One can be fairly confident that the cut point is beyond the last observed data point from the epidemiology studies (e.g., beyond 236 /100ml and a 10/1000 risk for *E coli* in fresh water. However, since the epidemiology data is illness based and the dosing trial data is pathogen-specific, it is difficult to estimate this cut-off using the higher doses from the dosing trials. One approach may be to use sensitivity studies to looking at highly infectious organisms and less infectious organisms.

Charles McGee

Unless there were measurements of water quality and illness that would allow the linear regression to be defined beyond the 1% level, the answer to this question is no. However, Dr. Kim and I suggest that data could be used from other studies with a similar experimental design (such as the Santa Monica Bay) to supplement EPA's original regression calculations. Dr. Kim has already examined the Santa Monica Bay data and verified that a similar illness relationship holds up to the 1% risk level. Having proven that, then it would be defensible to use data from that study or others to examine illness rates beyond the original 1%.

Mark Sobsey

It is scientifically defensible in principle to perform downward extrapolations to lower levels of risk on the basis of data for microbial dose and health effects response. This is done often in quantitative microbial risk assessment and in other health effects analyses. However, the scientific validity of doing this for the U.S. EPA data of this study and by the downward extrapolation method employed is not scientifically defensible. There is simply too much variability and uncertainty in the data to justify this downward extrapolation. Furthermore, the simple log-linear regression model used for this downward extrapolation is not adequately explained or justified and it is not compared to other more robust and scientifically valid downward extrapolation models for such data. Furthermore, the analyses does not report any sensitivity analyses that would indicate to what extent the output results would changes due to changes in microbial water quality or changes in health effects outcome (illness rates). Nearly all of the potential sources of bias that would be factors influencing the results were not addressed in either the collection or presentation of the data or by accounting or controlling for them in the analyses performed.

3. How much further could one extrapolate and what would be the rationale for extrapolating further?

Joseph Eisenberg

Based on my answer to Question 2, I would not recommend extrapolation. The only reason to extrapolate beyond the data would be to provide water quality guidelines for greater than a 10/1000 risk level for freshwater exposures. However the reason for using risk levels greater than 10/1000 depends on how the acceptable level of risk is set, which is beyond the scope of this document.

That being said, based on some of the supporting documents sent to me, it may be defensible to provide guidelines above 10/1000 for certain situations. For example, Figure 7 (Cabelli 1983, EPA 600/1-80-031 [*Health Effects Criteria for Marine Recreational Waters*], p. 36) suggests that the data would allow guidance for 12–13 per 1000 illness for enterococci in fresh waters. Likewise, Figures 1 and 2 (Dufour 1984, EPA 600/1-84-004 [*Health Effects Criteria for Fresh Recreational Waters*], p. 26) suggest that the data would allow guidance to just below 30 per 1000 illnesses for enterococci in marine waters.

Charles McGee

The answer to this question will be limited by water quality measured in other studies. The extrapolation has not been done yet, but Dr. Kim would like the opportunity to do so using the Santa Monica Bay data.

Mark Sobsey

As indicated in the response to Question 2, it is scientifically defensible to perform downward extrapolations to ranges of dose-response that are well below the levels of the observable data range. These extrapolations can be by as much as several orders of magnitude in some quantitative microbial dose-response and risk assessment analyses. However, as indicated in the response to Question 2, such downward extrapolations are not justified for these data. This is because of the limitations in the quantity and quality of the data, the failure to account for or control for bias in either the data collection and data analyses, and the limitations of the downward extrapolation analyses. Specifically only one simple log-linear regression model was used, there was an inadequate effort to address variability and uncertain and there is a lack of any sensitivity analyses.

V. SPECIFIC COMMENTS

Mark Sobsey

Analysis and Discussion

Limitations of Study Data and Analyses

This risk analysis is inadequate because the quantity and quality of the data used are inadequate and because the analytical approach and execution also are inadequate. Overall, the U.S. EPA data and the analytic approach are inadequate to address major sources of bias that can influence the data quality and the analytic approaches from which to estimate health risks in relation to water quality.

A major limitation of the proposed "*Implementation Guidance for Ambient Water Quality Criteria for Bacteria*" is the overall quality of the microbiological and human health effects data used for the analysis that provides the basis for the criteria. More specifically, the data come from studies at only three marine beach locations and only two freshwater beach locations. These studies did not adequately address other and more diversified sources of fecal contamination, such as more highly treated sewage effluents in which the ratios of fecal indicator bacteria to pathogens may be different than those at the few beaches studied. These few studies and study sites also do not adequately represent some other sources of fecal contamination that can impact bathing water and carry pathogens, such as non-point sources of human fecal contamination (e.g., septic tank-soil absorption systems and waste discharges from boats in nearby marinas) or non-human fecal contamination sources, such as waterfowl and animal agricultural waste. In many beach locations these other sources are the major sources of fecal contamination and they may results in different relationships among fecal indicator bacteria, pathogens and attendant human health risks.

Additionally, the important advances in statistical methods that have been made in the last two decades and are now widely used in health effects research for dose-response relationships and in quantitative microbial risk assessment were not applied in this study. Advanced regression and multivariate analyses methods were not applied in these studies and they should have been. Such advanced analytical methods are also much better for addressing variability and uncertainty and in controlling for bias.

There are many sources of bias in these studies and these sources of bias were not adequately addressed or controlled for in the U.S. EPA studies. These sources of bias, most of which apply to the U.S. EPA studies, are summarized in Table 1.

TABLE 1. Types of Biases Potentially Encountered in Recreational Water Quality Health Effects Studies and Their Potential Effects

Type of Bias	Description
Use of indicator microbes to assess water quality of exposure	Temporal and spatial indicator variation is substantial and difficult to relate to individual bathers (Fleisher, 1990), unless study design is experimental (Kay et al., 1994; Fleisher et al., 1996a). This is a limitation of the U.S. EPA data. Limited precision of methods for counting indicator organisms, causing measurement error (Fleisher, 1990; Fleisher et al., 1993); bacterial indicators may not be representative of viruses, which may be important etiological agents of swimming associated gastrointestinal illness. This is another limitation of the U.S. EPA data.
Use of seasonal means to assess water quality	Some studies use seasonal or other collapsed or grouped means and not daily measurements of indicator organisms to characterize individual exposure, thus adding substantial inaccuracy. This is a limitation of the U.S. EPA analysis.
Assessment of exposure pathway	Certain studies do not account for the potential infection pathway to definite exposure, e.g., mainly head immersion or ingestion of water for gastrointestinal symptoms. Difficulties in exposure recall further increase inaccuracy of individual exposure. These were limitations in the U.S. EPA studies.
Non-control for confounders	Non-control for confounders (e.g., food and drink intake, age, sex, history of certain diseases, drug use, personal contact, additional bathing, sun, socio-economic factors, etc.), may influence the observed association. These were limitations in the U.S. EPA studies.
Selection of un-representative study population	Results reported for certain study populations (e.g., limited age groups or from regions with certain endemicities) are *a priori* not directly transferable to populations with other characteristics. This was a limitation of the U.S. EPA studies.
Self-reporting of symptoms	Most observational studies relied on self-reporting of symptoms by study populations. Validation of symptoms by medical examination (Kay et al., 1994; Fleisher et al., 1996a) would reduce potential bias. External factors, such as media or publicity, may have influenced self-reporting. This was a limitation of the U.S. EPA study.
Response rate	Response rates were >70% in all, and >80% in most, studies. Differential reporting, e.g., higher response among participants experiencing symptoms, would probably not have major consequences.

Type of Bias	Description
Recruitment method	Recruitment methods were to approach persons on beaches in almost all observational studies and by advertisement for randomized controlled studies.
Interviewer effect	Differences in methodology of data collection among interviewers may influence the study results.

(Adapted from Pruss, 1998; Stavros and Langford, 2002; WHO, 2001)

Data Quality and Quantity Used for the Risk Analysis

The quality and quantity of the data used for the analysis are too limited and inadequate to provide reliable national estimates of the relationships between human exposure to pathogens (dose) as estimated by measuring fecal indicator bacteria such as *E. coli* and enterococci and human responses (health effects) in bathers. More and better data are available from numerous studies, some in the USA and some in other countries, and they should have been used for these analyses. In addition, even for the single set of data that was analyzed, additional analytical methods should have been employed to provide potentially better estimates of the relationships between bacterial quality of water and human health effects in bathers.

Limited data for a few geographic locations and beaches were used for the analyses. For marine water beaches only three different geographic locations were used as study sites, New York City, Lake Pontchartrain, LA, and Boston, MA. New York City had two beaches, one relatively polluted and one relatively unpolluted. The Lake Pontchartrain study had two beaches, both of which were impacted by less defined sources of fecal contamination than the point sources found at other study locations. Fecal contamination was believed to be caused by stormwater discharges reaching the beach via canals and bayous which empty into the Lake and elevated fecal indicator bacteria levels were observed in association with storm events. In Boston Harbor two beaches were studied. The pollution sources impacting these beaches were not as well defined as those at the New York City beaches. One of the beaches had fecal indicator bacteria levels about 1 order of magnitude higher than the other. While the 3 marine beach locations show some diversity and have 2 or more different beach sites for study, there are considerably more marine beaches with greater diversity of data that have been studied for relationships of bacteriological water quality and health effects in swimmers than are represented here.

For freshwater beaches, only two geographic locations, Erie, PA, and Keystone Lake, near Tulsa, OK, were used to represent all freshwater beaches of the entire country. At each geographic location, only two beaches were used, one of which was closer to a point source discharge of sewage effluent and the other of which was more remote from the sewage effluent source. The distances from the sewage effluent source to the beaches were not consistent from one geographical study site to the other, the degree of dilution of the sewage in the ambient water was not reported, and the quality of the sewage effluent varied. In one location it was chlorinated secondary effluent with unreported chlorine doses, unreported contact times and unreported residual concentrations of combined and free chlorine levels in the effluent when discharged. In the other location, the effluent differed from one study year to the other. Initially it was undisinfected effluent from two "full retention" lagoons of unreported type (anaerobic, facultative or aerobic), retention time and operating conditions (e.g., were the two lagoons operated in series or in parallel?). In the second year the effluent was treated in a lagoon of unreported type and retention time, followed by aeration (of unreported duration and

inadequate process description) followed by chlorination with unreported chlorine dose, unreported contact time and unreported residual levels of free or combined chlorine in the discharged effluent.

The data for the two geographic locations of freshwater beaches had considerable variability and uncertainty. For example in the 1979 study year at Lake Erie, the indicator densities were unexpectedly low at both beaches, in 1980 they were high and occasionally extremely high, and in 1982 they were moderately high relative to those observed in 1979.

At the Lake Erie beaches, the 1980 data do not reflect bacterial indicator densities consistent with proximity of the beach to the fecal waste source. In particular, the *E. coli* densities are higher at the beach more distant from the pollution source. The investigators suggest that these inconsistent results may have been caused by heavy rains which occurred in the four days before the start of the beach study trials. In a four-day period, 8.15 inches of rain was measured, which caused the lake elevation to rise and the turbidity to increase. The effect of these unusual events on the swimmer illness rates is unknown. These observations indicate a highly variable and not necessarily representative set of conditions at these two study locations that call into question their national representativeness of freshwater beaches.

There are also important limitation in the quality, quantity and representativeness of the bacteriological data for water quality. At each geographic location, only a few bacteriological measurements were made on any given day of observations. At each marine beach, samples were taken at 2-3 locations in chest deep water (location) and a given day of exposure, with only a few (3-4) samples collected between the hours of 11 AM and 5 PM (time of exposure). The actual numbers of water samples taken per location, site and study day were not specified for the studies at the freshwater beaches. However, it is said that the experimental design and approach was similar to that for the marine beaches studies.

It is unlikely that estimating the bacteriological quality of water based on a relatively small number of samples collected only in chest-deep water is representative of the exposure of all bathers. Children and many other people never venture into chest deep water and their exposures are likely to be better represented by water that is ankle-deep, knee-deep or waist deep. Some serious swimmers are also more likely to be exposed to water beyond the chest-deep area. In some subsequent studies on bathing water quality and health done by other investigators, water samples were collected from a grid of sample sites representing different depths and they were collected are more frequent intervals. Such sample provides better estimates of the bacteriological quality of the water at specific locations and times that can be referenced or related to the exposure of specific bathers who bathed at specific locations for specific time periods.

The best of the studies that related exposure to recreational bathing water of measured bacteriological quality to human health effects in those exposed by such bathing were randomized controlled trials in which subjected were recruited and randomly assigned to a bathing group or a non-bathing group. Bathers were asked to spend specific times in the water (10 minutes) and asked to immerse their heads at least three times (Kay et al., 1994; 2001; Fleisher et al., 1998). Exposure measurements of bacteriological quality of the water were made at the time and location of exposure and at three different depths, specifically surf, mid and chest.

Other Source of Relevant Data Not Used in this Analysis

Since these few early studies by the U.S. EPA, there have been numerous studies of water quality at bathing beaches in relation to human health effects in swimmers. The U.S. EPA should have used the marine water and freshwater data from these other studies in its analyses and risk assessment. In a review article published in 1998, Pruss reported on the findings of 4 additional studies conducted at freshwater beaches in the UK, France and Canada and 14 marine water beaches in countries in North America, Europe, the Middle East, Asia, the Western Pacific (Australia and New Zealand) and South Africa. More recently, Wade et al. (2003) conducted a systematic review and meta-analysis of a total of 15 marine water studies and 8 freshwater studies and on the relationships of bacterial water quality and health, besides the studies done by the U.S. EPA. These more recent studies provide a more diversified and more representative database than the few studies used by the U.S. EPA. These more recent analyses reveal wider ranges of bacterial densities than studied by the U.S. EPA and document different dose-response relationships for bacterial densities and human health effects than those observed by the U.S. EPA in its studies. In the study by Wade et al. (2003) the analyses of the data for freshwater beaches showed elevated relative risks of swimming-associated illness both above and below the U.S. EPA guideline value for enterococci. For *E. coli*, studies below the U.S. EPA guideline value were not associated with increased illness, while exposures above the U.S. EPA guideline value were. These findings suggest that the current U.S. EPA guideline values for the two different indicators provide different and inconsistent levels of bather protection from swimming-associated illness in freshwater beaches. Therefore, the two different indicators and their associated guideline values are not interchangeable in terms of their levels of protection. Based on the U.S. EPA's analysis of its own data for freshwater studies, these two indicators are considered interchangeable and give equivalent levels of protection. For the analyses of the marine water beaches, Wade et al. (2003) found that enterococci were the indicator that most strongly predicted increased health risks. By categorical analysis, the relative health risks did not continue to increase in studies with bacterial densities greater than 104 cfu/100 ml. This indicates a potential threshold effect for risk of GI illness. By weighted regression analysis there was an association between enterococci density and the natural log of the relative risk of health effects. The relative risk for GI illness increased 1.3 times for every log10 increase in enterococci density in water. In relation to the current U.S. EPA guideline for enterococci in marine waters, summary relative risks for GI illness below the U.S. EPA guideline value were lower and not statistically significant. Relative risks for GI illness above the U.S. EPA guideline value were elevated and statistically significant.

Analytical Approach and Execution

The approach to the analyses of the microbiological data is limited and probably flawed. More and better approaches to the types of data gathered and analyzed and the type of analyses of the data should have been conducted. The U.S. EPA should have attempted to provide better estimates of the concentrations of bacteria in the water to which people were exposed, particularly with respect to the spatial and temporal relationships of the exposures. The bacterial data for each location were analyzed by grouping them by location and then season and computing a geometric mean concentration of bacteria. Apparently, only a few measures of bacteria concentration were made for a given beach, with samples taken at 2-3 locations n chest deep water (location) and a given day of exposure, with only a few (3-4) samples collected between the hours of 11 AM and 5 PM (time of exposure). Geometric mean concentrations were computed for this exposure location and day and then used in a linear regression analysis to determine relationships between illness rates and average

water quality. The resulting linear regression model was then used to derive standard deviations for the dose-response relationship.

The approach used by the U.S. EPA in which all data were combined for the development of the distribution of bacterial concentrations that was then used for comparisons with human health effects has serious drawbacks and limitations. This is because the standard deviation of the probability density function or distribution affects the probability of exposure to polluted water and thus the risk of illness. The use of a single distribution and resulting parameter value to define a guideline value across all waters does not adequately address local variability. Local variations in standard deviation will mean that risks of illness will vary even though the same guideline value is in place. Simply stated, combining the data into a single distribution does not address local variability in the distribution of bacteria density on a site-specific basis and the relationships of these bacterial densities to local heath risks.

In addition, it is not clear the U.S. EPA used robust criteria to determine if bacterial concentrations could be legitimately log10-transformed to create the log10 distribution that was used in the dose-response analyses. Statistical tests should have been done to test for normality or to determine if the hypothesis of normality can be rejected. Such analyses should have been done for the data from individual beaches and study sites as well as on the combined data. The need to test for the normality of the data on a site-specific basis is because of the need to link bacterial densities with specific exposures that result in health effects.

Apparently, the U.S. EPA did not consider alternative analytical strategies for these data that do not rely on the use of the data taken directly from a log10-transformed distribution. One possible alternative approach in situations where normality (or log-normality) is violated is to use a Bootstrapping procedure. In this case a Monte Carlo style procedure is applied to bootstrapped samples of the actual empirical distributions of bacterial concentration, rather than the parametrically generated distribution. The bootstrapping procedure draws a large number (say 1000) of "resamples", of size equal to the original sample, from this original sample randomly with replacement. No such alternative analytical approach was attempted by the U.S. EPA.

The basis for defining the different risk levels as upper percentiles (e.g., 75^{th}, 82^{nd}, 90^{th} and 95th) is poorly documented and justified and the basis for focusing only the human health risks in the range of 0.8 to 1% also is poorly documented and inadequately explained or justified. These standard deviation estimates were used to consider various upper percentiles (75^{th}, 82^{d}, 90^{th} and 95%) for which were calculate various upper percentiles values of allowable bacterial density per 100 ml corresponding to different health risk levels in the range of 0.8 to 1%. The flaws and weakness of this approach are that the data were transformed to log10 values and these log10-transformed values were used in the analyses and the analysis of the log-transformed data used only a simple linear regression model to examine the relationship between indicator density and human health risk of bathers. No other forms of the data were used for analyses (such as arithmetic forms of the data with no transformations) and no other models were applied to the data (such as Beta-Poisson, two-population or other dose-response models now widely used for quantitative microbial risk assessment).

In the study of Kay et al. (1994) an epidemiological relationship described the excess risk of illness from exposure to water containing fecal indicator bacteria. This relationship was best as a dose-response relationship linking water quality exposure (x), indexed by the fecal streptococci density at chest depth water, and the excess probability of gastroenteritis (y) is given by the following (for exposures between 32 and 158 fecal streptococci per 100ml):

$$y_{(x=32:158)} = \frac{1}{1+e^{-m}} - p_{32}$$

where, *m* is the natural logarithm of the odds of getting gastroenteritis from bathing, derived from the logistic regression equation:

$$m = 0.20102\sqrt{x-32} - 2.3561$$

and the term *p32* is the probability of gastroenteritis where x = 32 cfu per 100 ml *(*p=0.0866) and adjusts the relationship to reflect excess rather than absolute probability of illness relative to those who do not bathe.

In addition to the consideration of the dose-response model used, there are valid reasons to believe that the form of microbiological data used, specifically the log10-transformed data of the U.S. EPA, is likely to underestimate exposure in dose-response analyses and that the liner regression model applied to these data is likely to provide a downward extrapolation that is a less reliable portrayal of the actual dose-response relationship. This is not always the case, as was found in the studies by Kay et al., where the bacteriological data were best described by a log-normal probability density function. However, there are scientifically valid reasons to use arithmetic data for dose-response analyses and determine if this from of the data better describes the extremes of exposure and resulting health effects at both the low and high ends of the distribution of bacteria concentration.
Overall, there are serious concerns about the extent to which reliable downward extrapolations can be made from the U.S. EPA data, given all of the potential sources of bias, the use of log10-transformed data rather than arithmetic data, the use of only a simple log-linear regression model, and the rather shallow slope of the dose-response relationship in the human health effects response range of interest (corresponding to 0.8 to 1.0% risk).

In addition, there was very inadequate treatment of variability and uncertainty in the analyses for either bacteriological quality of the water, the extent of bather exposure and the temporal and spatial relationships between exposure and resulting health effects in the exposed. These deficiencies make it inappropriate to attempt to do downward extrapolations of the water quality (bacterial concentration) - health effects) dose-response) relationships in the data range of interest for freshwaters or in general. Developing water quality criteria and regulatory guidelines based on such analyses can not be supported or justified in the opinion of this reviewer.

References Cited

Fleisher JM, Kay D, Wyer MD, A.F. Godfree Estimates of the severity of illnesses associated with bathing in marine recreational waters contaminated with domestic sewage. Int J Epidemiol. 1998 Aug;27(4):722-6.

Kay D, Fleisher JM, Salmon RL, Jones F, Wyer MD, Godfree AF, Zelenauch-Jacquotte Z, R. Shore (1994) Predicting likelihood of gastroenteritis from sea bathing: results from randomised exposure. Lancet. 1994 Oct 1;344(8927):905-9.

Kay, D., J. Fleisher, M.D. Wyer, and R.I. Salmon (2001) Re-analysis of the Seabathing Data from the UK Randomised Trials. A Report to DETR. Aberystwyth, University of Wales, Centre for Research into the Environment and health, 17 pages.

Pruss, A. (1998) Review of epidemiological studies on health effects from exposure to recreational water. International Journal of Epidemiology 27(1): 1-9.

Stavros, G. and I.H. Langford (2002) Coastal Bathing Water Quality and Human Health Risks: A Review of Legislation, Policy and Epidemiology, with an Assessment of Current UK Water Quality, Proposed Standards, and Disease Burden in England and Wales. CSERGE Working Paper ECM 02-06. Centre for Social and Economic Research on the Global Environment, University of East Anglia, UK.

Wade TJ, Pai N, Eisenberg JN, Colford JM Jr. (2003) Do U.S. Environmental Protection Agency water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis. Environ Health Perspect. 2003 Jun;111(8):1102-9.

WHO (2001) Bathing Water Quality and Human Health: Faecal Pollution. Outcome of an Expert Consultation, Farnham, UK, April 2001. Co-sponsored by Department of the Environment, Transport and the Regions, United Kingdom. WHO/SDE/WSH/01.2. Geneva: World Health Organization.

APPENDIX A

Joseph Eisenberg

Peer review of the EPA document"Implementation Guidance for Ambient Water Quality Criteria for Bacteria"

For

Versar

Reviewed by Joe Eisenberg
12/20/03

Overview

The predicted risk level (risk of GI illness per 1000 swimmers) for a given water quality (geometric mean density of an indicator organism) is based on a linear regression model. This model was estimated using beach study data (e.g., 9 data points for *E coli* in fresh waters). In addition to providing point risk estimates for geometric mean values of indicator densities, the regression model also allows for the calculation of 75, 82, 90, and 95% confidence limits, which were used to fill in the percentile values in Table 1-1 and 1-2.

Comments on text

I have some questions on the use of geometric means for estimating the dose of exposure. It is the arithmetic mean that provides the appropriate average exposure over time. The geometric mean, which is a better estimate of the median, will tend to underestimate the average level of exposure.

It would be nice to see Figures 1.3 and 1.4 for enterococci in fresh and marine waters.

1. Given the constraints of the data available, is the risk analysis in the *Implementation Guidance for Ambient Water Quality Criteria for Bacteria* appropriate?

   The guidelines are appropriate as written precisely because they do not go beyond the limits of the data. See answers to 2 and 3 for further clarification.

2. Is it scientifically defensible to extrapolate the relationship (in terms of linear regression or other quantitative means) between bacterial indicator density and illness rate for fresh waters beyond the 1% risk level?

   No. The linear regression model is a data driven model; i.e., there is no mechanistic rationale for the model structure. The predictions beyond the data, therefore, are not reliable or defensible.

   One potential way to extrapolate is to assume the sigmoidal dose-response relationship, based on data from dosing trials. These trials have been conducted for various pathogens and administer higher doses than observed in epidemiology studies. Given these trial data, one

can assume that the relationship is linear at low doses, increases exponentially at higher doses, and eventually saturates at yet higher doses. This is basically what is said in the document under review, and is illustrated in Figure 1.2. The problem becomes where to place the cut point. One can be fairly confident that the cut point is beyond the last observed data point from the epidemiology studies (e.g., beyond 236 /100ml and a 10/1000 risk for *E coli* in fresh water. However, since the epidemiology data is illness based and the dosing trial data is pathogen-specific, it is difficult to estimate this cut-off using the higher doses from the dosing trials. One approach may be to use sensitivity studies to looking at highly infectious organisms and less infectious organisms.

3. How much further could one extrapolate and what would be the rationale for extrapolating further?

   Based on my answer to Question 2, I would not recommend extrapolation. The only reason to extrapolate beyond the data would be to provide water quality guidelines for greater than a 10/1000 risk level for freshwater exposures. However the reason for using risk levels greater than 10/1000 depends on how the acceptable level of risk is set, which is beyond the scope of this document.

   That being said, based on some of the supporting documents sent to me, it may be defensible to provide guidelines above 10/1000 for certain situations. For example, Figure 7 (Cabelli 1983, EPA 600/1-84-004, p36) suggests that the data would allow guidance for 12 – 13 per 1000 illness for enterococci in fresh waters. Likewise, Figure 1 and 2 (Dufour 1984, EPA 600/1-84-004, p26) suggest that the data would allow guidance to just below 30 per 1000 illnesses for enterococci in marine waters.

APPENDIX B

Charles McGee

Diane S. Sinkowski
Environmental Engineer
Exposure/Risk Assessment Division
Versar, Inc.
6850 Versar Center
Springfield, VA 22151

Dear Ms. Sinknwski,

In regards to: Work Assignment #1-11; Peer Review of Epidemiological Data from EPA Bacteriological Studies

The following are my responses to the three questions raised in the above work assignment:

1. Given the constraints of the data available, is the risk analysis in the *Implementation Guidance for Ambient Water Quality Criteria for Bacteria* appropriate?

   Response: For many years, this analysis has been the subject of much debate. However, the experimental design, the quality of the data gathered and duplication of the results by other researchers has made the risk analysis put forth in the guidance defensible.

   A current challenge to the original research upon which the risk analysis is based is whether the spatial and temporal variability of the beach water quality was captured in the experimental design. In any study, the strength of the relationships between two variables is dependent on the precision of the measurement of those variables. EPA's own EMPACT study and research on recreational water contamination carried out on the west coast has demonstrated the significance of this variability. In preparing my answer to this question, I reviewed some of the original EPA publications, and I was convinced that the study design adequately addressed this concern.

   A second issue that should be addressed is how measurement error was taken into account in the calculation of risks criteria for indicator bacteria densities (see Table 4 in EPA440/5-84-002). Specifically, the question was whether water quality variation can be better explained by systematic patterns or is variability simply the result of random error. Dr. Joon Ha Kim and I had considered this point in a journal article evaluating error associated with California's marine water quality monitoring and public notification procedures. Upon submission of this article to the journal, one of the article's reviewer's comments triggered a rewrite of the article. The concepts in that rewrite not only apply to the expression of error in the formula in Table 4 but to all three questions posed in this peer review. Because of time constraints, I could not contribute to the rewrite and had my name removed as an author. I have spoken with Dr. Kim, and he has agreed to allow you to contact him for an advance copy of his new article.

He would also like to suggest some other data analyses of the EPA and Santa Monica Bay data. Dr. Kim's address is:

Joon Ha Kim, Ph.D.

944A Engineering Tower

Chemical Engineering & Materials Science

University of California, Irvine

Irvine, California 92697-2575

Phone number 949-824-7754

2. Is it scientifically defensible to extrapolate the relationship (in terms of linear regression or other quantitative means) between bacterial indicator density and illness rate for fresh waters beyond the 1% risk level?

Response: Unless there were measurements of water quality and illness that would allow the linear regression to be defined beyond the 1% level, the answer to this question is no. However, Dr. Kim and I suggest that data could be used from other studies with a similar experimental design (such as the Santa Monica Bay) to supplement EPA's original regression calculations. Dr. Kim has already examined the Santa Monica Bay data and verified that a similar illness relationship holds up to the 1% risk level. Having proven that, then it would be defensible to use data from that study or others to examine illness rates beyond the original 1%.

3. How much further could one extrapolate and what would be the rational for extrapolating further?

Response: The answer to this question will be limited by water quality measured in other studies. The extrapolation has not been done yet, but Dr. Kim would like the opportunity to do so using the Santa Monica Bay data.

Thank you for the opportunity to comment on these questions, but my best advice would be to contact Dr. Kim to actually perform the statistical analysis that could further strengthen the underpinnings of the guidance document.

Best regards,
Charles D. McGee
Laboratory Supervisor
Orange County Sanitation District

APPENDIX C

Mark Sobsey

Peer Review of Epidemiological Data from EPA Bacteriological Studies

Mark D. Sobsey
University of North Carolina
CB# 7431, McGavran-Greenberg Hall, Room 4114a
Chapel Hill, NC 27599-7431

Responses to the three specific questions posed to the reviewers:

1. Given the constraints of the data available, is the risk analysis in the *Implementation Guidance for Ambient Water Quality Criteria for Bacteria*" appropriate?

Response: The answer is "no". The reasons for this answer will be given below in more detail.

2. Is it scientifically defensible to extrapolate the relationship (in terms of linear regression or other quantitative means) between bacterial indicator density and illness rate for fresh waters beyond the 1% risk level?

Response: It is scientifically defensible in principle to perform downward extrapolations to lower levels of risk on the basis of data for microbial dose and health effects response. This is done often in quantitative microbial risk assessment and in other health effects analyses. However, the scientific validity of doing this for the US EPA data of this study and by the downward extrapolation method employed is not scientifically defensible. There is simply too much variability and uncertainty in the data to justify this downward extrapolation. Furthermore, the simple log-linear regression model used for this downward extrapolation is not adequately explained or justified and it is not compared to other more robust and scientifically valid downward extrapolation models for such data. Furthermore, the analyses does not report any sensitivity analyses that would indicate to what extent the output results would changes due to changes in microbial water quality or changes in health effects outcome (illness rates). Nearly all of the potential sources of bias that would be factors influencing the results were not addressed in either the collection or presentation of the data or by accounting or controlling for them in the analyses performed.

3. How much further could one extrapolate and what would be the rationale for extrapolation further.

Response: As indicated in the response to question 2, it is scientifically defensible to perform downward extrapolations to ranges of dose-response that are well below the levels of the observable data range. These extrapolations can be by as much as several orders of magnitude in some quantitative microbial dose-response and risk assessment analyses. However, as indicated in the response to question 2, such downward extrapolations are not justified for these data. This is because of the limitations in the quantity and quality of the data, the failure to account for or control for bias in either the data collection and data analyses, and the limitations of the downward extrapolation analyses. Specifically only one simple log-linear regression model was used, there was an inadequate effort to address variability and uncertain and there is a lack of any sensitivity analyses.

ANALYSIS AND DISCUSSION

Limitations of Study Data and Analyses

This risk analysis is inadequate because the quantity and quality of the data used are inadequate and because the analytical approach and execution also are inadequate. Overall, the US EPA data and the analytic approach are inadequate to address major sources of bias that can influence the data quality and the

analytic approaches from which to estimate health risks in relation to water quality.

A major limitation of the proposed "Implementation Guidance for Ambient Water Quality Criteria for bacteria" is the overall quality of the microbiological and human health effects data used for the analysis that provides the basis for the criteria. More specifically, the data come from studies at only three marine beach locations and only two freshwater beach locations. These studies did not adequately address other and more diversified sources of fecal contamination, such as more highly treated sewage effluents in which the ratios of fecal indicator bacteria to pathogens may be different than those at the few beaches studied. These few studies and study sites also do not adequately represent some other sources of fecal contamination that can impact bathing water and carry pathogens, such as non-point sources of human fecal contamination (e.g., septic tank-soil absorption systems and waste discharges from boats in nearby marinas) or non-human fecal contamination sources, such as waterfowl and animal agricultural waste. In many beach locations these other sources are the major sources of fecal contamination and they may results in different relationships among fecal indicator bacteria, pathogens and attendant human health risks.

Additionally, the important advances in statistical methods that have been made in the last two decades and are now widely used in health effects research for dose-response relationships and in quantitative microbial risk assessment were not applied in this study. Advanced regression and multivariate analyses methods were not applied in these studies and they should have been. Such advanced analytical methods are also much better for addressing variability and uncertainty and in controlling for bias.

There are many sources of bias in these studies and these sources of bias were not adequately addressed or controlled for in the EPA studies. These sources of bias, most of which apply to the US EPA studies, are summarized in Table 1 below.

Type of Bias	Description
Use of indicator microbes to assess water quality of exposure	Temporal and spatial indicator variation is substantial and difficult to relate to individual bathers (Fleisher, 1990), unless study design is experimental (Kay et al., 1994; Fleisher et al., 1996a). This is a limitation of the US EPA data. Limited precision of methods for counting indicator organisms, causing measurement error (Fleisher, 1990; Fleisher et al., 1993); bacterial indicators may not be representative of viruses, which may be important etiological agents of swimming associated gastrointestinal illness. This is another limitation of the US EPA data.
Use of seasonal means to assess water quality	Some studies use seasonal or other collapsed or grouped means and not daily measurements of indicator organisms to characterize individual exposure, thus adding substantial inaccuracy. This is a limitation of the US EPA analysis.
Assessment of exposure pathway	Certain studies do not account for the potential infection pathway to definite exposure, e.g., mainly head immersion or ingestion of water for gastrointestinal symptoms. Difficulties in exposure recall further increase inaccuracy of individual exposure. These were limitations in the US EPA studies.
Non-control for confounders	Non-control for confounders (e.g., food and drink intake, age, sex, history of certain diseases, drug use, personal contact, additional bathing, sun, socio-economic factors, etc.), may influence the observed association. These were limitations in the US EPA studies.
Selection of	Results reported for certain study populations (e.g., limited

un-representative study population	age groups or from regions with certain endemicities) are *a priori* not directly transferable to populations with other characteristics. This was a limitation of the US EPA studies.
Self-reporting of symptoms	Most observational studies relied on self-reporting of symptoms by study populations. Validation of symptoms by medical examination (Kay et al., 1994; Fleisher et al., 1996a) would reduce potential bias. External factors, such as media or publicity, may have influenced self-reporting. This was a limitation of the US EPA study.
Response rate	Response rates were >70% in all, and >80% in most, studies. Differential reporting, e.g., higher response among participants experiencing symptoms, would probably not have major consequences.
Recruitment method	Recruitment methods were to approach persons on beaches in almost all observational studies and by advertisement for randomized controlled studies.
Interviewer effect	Differences in methodology of data collection among interviewers may influence the study results.

TABLE 1. Types of Biases Potentially Encountered in Recreational Water Quality Health Effects Studies and Their Potential Effects
(Adapted from Pruss, 1998; Stavros and Langford, 2002; WHO, 2001)

Data Quality and Quantity Used for the Risk Analysis

The quality and quantity of the data used for the analysis are too limited and inadequate to provide reliable national estimates of the relationships between human exposure to pathogens (dose) as estimated by measuring fecal indicator bacteria such as E. coli and enterococci and human responses (health effects) in bathers. More and better data are available from numerous studies, some in the USA and some in other countries, and they should have been used for these analyses. In addition, even for the single set of data that was analyzed, additional analytical methods should have been employed to provide potentially better estimates of the relationships between bacterial quality of water and human health effects in bathers.

Limited data for a few geographic locations and beaches were used for the analyses. For marine water beaches only three different geographic locations were used as study sites, New York City, Lake Pontchartrain, LA., and Boston, MA. New York City had two beaches, one relatively polluted and one relatively unpolluted. The Lake Pontchartrain study had two beaches, both of which were impacted by less defined sources of fecal contamination than the point sources found at other study locations. Fecal contamination was believed to be caused by stormwater discharges reaching the beach via canals and bayous which empty into the Lake and elevated fecal indicator bacteria levels were observed in association with storm events. In Boston harbor two beaches were studied. The pollution sources impacting these beaches were not as well defined as those at the New York City beaches. One of the beaches had fecal indicator bacteria levels about 1 order of magnitude higher than the other. While the 3 marine beach locations show some diversity and have 2 or more different beach sites for study, there are considerably more marine beaches with greater diversity of data that have been studied for relationships of bacteriological water quality and health effects in swimmers than are represented here.

For freshwater beaches, only two geographic locations, Erie, PA and Keystone Lake, near Tulsa, OK, were used to represent all freshwater beaches of the entire country (. At each geographic location, only two beaches were used, one of which was closer to a point source discharge of sewage effluent and the other of which was more remote from the sewage effluent source. The distances from the sewage effluent source to the beaches were not consistent from one geographical study

site to the other, the degree of dilution of the sewage in the ambient water was not reported, and the quality of the sewage effluent varied. In one location it was chlorinated secondary effluent with unreported chlorine doses, unreported contact times and unreported residual concentrations of combined and free chlorine levels in the effluent when discharged. In the other location, the effluent differed from one study year to the other. Initially it was undisinfected effluent from two "full retention" lagoons of unreported type (anaerobic, facultative or aerobic), retention time and operating conditions (e.g., were the two lagoons operated in series or in parallel?). In the second year the effluent was treated in a lagoon of unreported type and retention time, followed by aeration (of unreported duration and inadequate process description) followed by chlorination with unreported chlorine dose, unreported contact time and unreported residual levels of free or combined chlorine in the discharged effluent.

The data for the two geographic locations of freshwater beaches had considerable variability and uncertainty. For example in the 1979 study year at Lake Erie, the indicator densities were unexpectedly low at both beaches, in 1980 they were high and occasionally extremely high, and in 1982 they were moderately high relative to those observed in 1979.

At the Lake Erie beaches, the 1980 data do not reflect bacterial indicator densities consistent with proximity of the beach to the fecal waste source. In particular, the *E. coli* densities are higher at the beach more distant from the pollution source. The investigators suggest that these inconsistent results may have been caused by heavy rains which occurred in the four days before the start of the beach study trials. In a four-day period, 8.15 inches of rain was measured, which caused the lake elevation to rise and the turbidity to increase The effect of these unusual events on the swimmer illness rates is unknown. These observations indicate a highly variable and not necessarily representative set of conditions at these two study locations that call into question their national representativeness of freshwater beaches.

There are also important limitation in the quality, quantity and representativeness of the bacteriological data for water quality. At each geographic location, only a few bacteriological measurements were made on any given day of observations. At each marine beach, samples were taken at 2-3 locations in chest deep water (location) and a given day of exposure, with only a few (3-4) samples collected between the hours of 11 AM and 5 PM (time of exposure). The actual numbers of water samples taken per location, site and study day were not specified for the studies at the freshwater beaches. However, it is said that the experimental design and approach was similar to that for the marine beaches studies.

It is unlikely that estimating the bacteriological quality of water based on a relatively small number of samples collected only in chest-deep water is representative of the exposure of all bathers. Children and many other people never venture into chest deep water and their exposures are likely to be better represented by water that is ankle-deep, knee-deep or waist deep. Some serious swimmers are also more likely to be exposed to water beyond the chest-deep area. In some subsequent studies on bathing water quality and health done by other investigators, water samples were collected from a grid of sample sites representing different depths and they were collected are more frequent intervals. Such sample provides better estimates of the bacteriological quality of the water at specific locations and times that can be referenced or related to the exposure of specific bathers who bathed at specific locations for specific time periods.

The best of the studies that related exposure to recreational bathing water of measured bacteriological quality to human health effects in those exposed by such bathing were randomized controlled trials in which subjected were recruited and randomly assigned to a bathing group or a non-bathing group. Bathers were asked to spend specific times in the water (10 minutes) and asked to immerse their heads at least three times (Kay et al., 1994; 2001; Fleisher et al., 1998). Exposure measurements of bacteriological quality of the water were made at the

time and location of exposure and at three different depths, specifically surf, mid and chest.

Other Source of Relevant Data Not Used in this Analysis

Since these few early studies by the US EPA, there have been numerous studies of water quality at bathing beaches in relation to human health effects in swimmers. The US EPA should have used the marine water and freshwater data from these other studies in its analyses and risk assessment. In a review article published in 1998, Preuss reported on the findings of 4 additional studies conducted at freshwater beaches in the UK, France and Canada and 14 marine water beaches in countries in North America, Europe, the Middle East, Asia, the Western Pacific (Australia and New Zealand) and South Africa. More recently, Wade et al. (2003) conducted a systematic review and meta-analysis of a total of 15 marine water studies and 8 freshwater studies and on the relationships of bacterial water quality and health, besides the studies done by the US EPA. These more recent studies provide a more diversified and more representative database than the few studies used by the US EPA. These more recent analyses reveal wider ranges of bacterial densities than studied by the US EPA and document different dose-response relationships for bacterial densities and human health effects than those observed by the US EPA in its studies. In the study by Wade et al. (2003) the analyses of the data for freshwater beaches showed elevated relative risks of swimming-associated illness both above and below the US EPA guideline value for enterococci. For E. coli, studies below the US EPA guideline value were not associated with increased illness, while exposures above the US EPA guideline value were. These findings suggest that the current US EPA guideline values for the two different indicators provide different and inconsistent levels of bather protection from swimming-associated illness in freshwater beaches. Therefore, the two different indicators and their associated guideline values are not interchangeable in terms of their levels of protection. Based on the EPA's analysis of its own data for freshwater studies, these two indicators are considered interchangeable and give equivalent levels of protection. For the analyses of the marine water beaches, Wade et al. (2003) found that enterococci were the indicator that most strongly predicted increased health risks. By categorical analysis, the relative health risks did not continue to increase in studies with bacterial densities greater than 104 cfu/100 ml. This indicates a potential threshold effect for risk of GI illness. By weighted regression analysis there was an association between enterococci density and the natural log of the relative risk of health effects. The relative risk for GI illness increased 1.3 times for every log10 increase in enterococci density in water. In relation to the current EPA guideline for enterococci in marine waters, summary relative risks for GI illness below the EPA guideline value were lower and not statistically significant. Relative risks for GI illness above the EPA guideline value were elevated and statistically significant.

Analytical Approach and Execution

The approach to the analyses of the microbiological data is limited and probably flawed. More and better approaches to the types of data gathered and analyzed and the type of analyses of the data should have been conducted. The EPA should have attempted to provide better estimates of the concentrations of bacteria in the water to which people were exposed, particularly with respect to the spatial and temporal relationships of the exposures. The bacterial data for each location were analyzed by grouping them by location and then season and computing a geometric mean concentration of bacteria. Apparently, only a few measures of bacteria concentration were made for a given beach, with samples taken at 2-3 locations n chest deep water (location) and a given day of exposure, with only a few (3-4) samples collected between the hours of 11 AM and 5 PM (time of exposure). Geometric mean concentrations were computed for this exposure location and day and then used in a linear regression analysis to determine relationships between illness rates and average water quality. The resulting linear regression model was then used to derive standard deviations for the dose-response relationship.

The approach used by the US EPA in which all data were combined for the development of the distribution of bacterial concentrations that was then used for comparisons with human health effects has serious drawbacks and limitations. This is because the standard deviation of the probability density function or distribution affects the probability of exposure to polluted water and thus the risk of illness. The use of a single distribution and resulting parameter value to define a guideline value across all waters does not adequately address local variability. Local variations in standard deviation will mean that risks of illness will vary even though the same guideline value is in place. Simply stated, combining the data into a single distribution does not address local variability in the distribution of bacteria density on a site-specific basis and the relationships of these bacterial densities to local heath risks

In addition, it is not clear the US EPA used robust criteria to determine if bacterial concentrations could be legitimately log10 transformed to create the log10 distribution that was used in the dose-response analyses. Statistical tests should have been done to test for normality or to determine if the hypothesis of normality can be rejected. Such analyses should have been done for the data from individual beaches and study sites as well as on the combined data. The need to test for the normality of the data on a site-specific basis is because of the need to link bacterial densities with specific exposures that result in health effects.

Apparently, the US EPA did not consider alternative analytical strategies for these data that do not rely on the use of the data taken directly from a log10-transformed distribution. One possible alternative approach in situations where normality (or log-normality) is violated is to use a Bootstrapping procedure. In this case a Monte Carlo style procedure is applied to bootstrapped samples of the actual empirical distributions of bacterial concentration, rather than the parametrically generated distribution. The bootstrapping procedure draws a large number (say 1000) of "resamples", of size equal to the original sample, from this original sample randomly with replacement. No such alternative analytical approach was attempted by the use EPA.

The basis for defining the di fferent risk levels as upper percentiles (e.g., 75th, 82nd, 90th and 95th) is poorly documented and justified and the basis for focusing only the human health risks in the range of 0.8 to 1% also is poorly documented and inadequately explained or justified.
These standard deviation estimates were used to consider various upper percentiles (75th, 82nd, 90th and 95%) for which were calculate various upper percentiles values of allowable bacterial density per 100 ml corresponding to different health risk levels in the range of 0.8 to 1%. The flaws and weakness of this approach are that the data were transformed to log10 values and these log10 transformed values were used in the analyses and the analysis of the log-transformed data used only a simple linear regression model to examine the relationship between indicator density and human health risk of bathers. No other forms of the data were used for analyses (such as arithmetic forms of the data with no transformations) and no other models were applied to the data (such as Beta-Poisson, two-population or other dose-response models now widely used for quantitative microbial risk assessment).

In the study of Kay et al. (1994) an epidemiological relationship described the excess risk of illness from exposure to water containing fecal indicator bacteria. This relationship was best as a dose-response relationship linking water quality exposure (x), indexed by the fecal streptococci density at chest depth water, and the excess probability of gastroenteritis (y) is given by the following (for exposures between 32 and 158 fecal streptococci per 100ml):

$$y_{(x=32:158)} = \frac{1}{1+e^{-m}} - p_{32}$$

where, *m* is the natural logarithm of the odds of getting gastroenteritis from bathing, derived from the logistic regression equation:

$$m = 0.20102\sqrt{x - 32} - 2.3561$$

and the term *p32* is the probability of gastroenteritis where x = 32 cfu per 100ml (p=0.0866) and adjusts the relationship to reflect excess rather than absolute probability of illness relative to those who do not bathe.

In addition to the consideration of the dose-response model used, there are valid reasons to believe that the form of microbiological data used, specifically the log10—transformed data of the US EPA, is likely to underestimate exposure in dose-response analyses and that the liner regression model applied to these data is likely to provide a downward extrapolation that is a less reliable portrayal of the actual dose-response relationship. This is not always the case, as was found in the studies by Kay et al., where the bacteriological data were best described by a log-normal probability density function. However, there are scientifically valid reasons to use arithmetic data for dose-response analyses and determine if this from of the data better describes the extremes of exposure and resulting health effects at both the low and high ends of the distribution of bacteria concentration.

Overall, there are serious concerns about the extent to which reliable downward extrapolations can be made from the US EPA data, given all of the potential sources of bias, the use of loh10-transformed data rather than arithmetic data, the use of only a simple log-linear regression model, and the rather shallow slope of the dose-response relationship in the human health effects response range of interest (corresponding to 0.8 to 1.0% risk).

In addition, there was very inadequate treatment of variability and uncertainty in the analyses for either bacteriological quality of the water, the extent of bather exposure and the temporal and spatial relationships between exposure and resulting health effects in the exposed. These deficiencies make it inappropriate to attempt to do downward extrapolations of the water quality (bacterial concentration)-health effects)dose-response) relationships in the data range of interest for freshwaters or in general. Developing water quality criteria and regulatory guidelines based on such analyses can not be supported or justified in the opinion of this reviewer.

References Cited

Fleisher JM, Kay D, Wyer MD, A.F. Godfree Estimates of the severity of illnesses associated with bathing in marine recreational waters contaminated with domestic sewage. Int J Epidemiol. 1998 Aug;27(4):722-6.

Kay D, Fleisher JM, Salmon RL, Jones F, Wyer MD, Godfree AF, Zelenauch-Jacquotte Z, R. Shore (1994) Predicting likelihood of gastroenteritis from sea bathing: results from randomised exposure. Lancet. 1994 Oct 1;344(8927):905-9.

Kay, D., J. Fleisher, M.D. Wyer, and R.I. Salmon (2001) Re-analysis of the Seabathing Data from the UK Randomised Trials. A Report to DETR. Aberystwyth, University of Wales, Centre for Research into the Environment and health, 17 pages.

Pruss, A. (1998) Review of epidemiological studies on health effects from exposure to recreational water. International Journal of Epidemiology 27(1): 1-9.

Stavros, G. and I.H. Langford (2002) Coastal Bathing Water Quality and Human Health Risks: A Review of Legislation, Policy and Epidemiology, with an Assessment of Current UK Water Quality, Proposed Standards, and Disease Burden in England and Wales. CSERGE Working Paper ECM 02-06. Centre for Social and Economic Research on the Global Environment, University of East Anglia, UK.

Wade TJ, Pai N, Eisenberg JN, Colford JM Jr. (2003) Do U.S. Environmental Protection Agency water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis. Environ Health Perspect. 2003 Jun;111(8):1102-9.

WHO (2001) Bathing Water Quality and Human Health: Faecal Pollution. Outcome of an Expert Consultation, Farnham, UK, April 2001. Co-sponsored by Department of the Environment, Transport and the Regions, United Kingdom. WHO/SDE/WSH/01.2. Geneva: World Health Organization.