Water Quality Centre Publication No. 18

BACKGROUND NOTES FOR THE DEVELOPMENT OF GUIDELINES FOR MICROBIOLOGICAL RECEIVING WATER STANDARDS FOR NEW ZEALAND

WATER QUALITY CENTRE

Environmental Research and Consulting

Water Quality Centre Publication No. 18

BACKGROUND NOTES FOR THE DEVELOPMENT OF GUIDELINES FOR MICROBIOLOGICAL RECEIVING WATER STANDARDS FOR NEW ZEALAND

by

Graham B McBride
Water Quality Centre
Department of Scientific and Industrial Research
P.O. Box 11-115, Hamilton, New Zealand

April 1990

Cataloguing-in-publication

MCBRIDE, Graham Burnley
Background notes for the development of guidelines for microbiological receiving water standards for New Zealand / by Graham B. McBride - Hamilton, NZ: Water Quality Centre, DSIR, 1990. (Water Quality Centre publication; 18.
ISSN 0112-689X. ISBN 0-477-02589-7)

I. Title II. Water Quality Centre (Hamilton, NZ). III. Series

ERRATA

Water Quality Centre Publication No. 18

BACKGROUND NOTES FOR THE DEVELOPMENT OF GUIDELINES FOR MICROBIOLOGICAL RECEIVING WATER STANDARDS FOR NEW ZEALAND

Please make the following corrections to the References:

p. 16, in Cabelli, V.J., 1983, change the report number to **EPA-600/1-80-031.**

p. 16, in Dufour, A.P., 1984, change the report number to **EPA-600/1-84-004.**

p. 18, in USEPA 1986b, change the report number to **EPA440/5-84-002.**

G B McBride, 29/5/90.

BACKGROUND NOTES FOR THE DEVELOPMENT OF GUIDELINES FOR MICROBIOLOGICAL RECEIVING WATER STANDARDS FOR NEW ZEALAND

ABSTRACT

Proposals in the Resource Management Bill introduced to the New Zealand Parliament in late 1989 give regional government the function of preparing "regional plans". Such plans can include the setting of microbiological standards for inland and coastal waters, and so have the ability to constrain land developments and the location and treatment of waste discharges. The microbiological standards given in the Bill are in narrative form, requiring that undesirable outcomes do not occur as a result of microbiological contamination. So when plans come to be drawn up, as much as possible the narrative statement will have to be translated into an explicit numerical form, identifying what should be measured, and stating a compliance rule. If such explicit standards are to be meaningful, considerable care will have to be used in this translation. Indeed, there is a need to develop guidelines for New Zealand to assist in such translations, and so maximise the usefulness of the plans, and minimise the time and cost spent in disputes and semi-judicial hearings.

This report represents the first step on the way to the production of guidelines. First, it traces the history of the development and application of such microbiological water standards as New Zealand has had, or have been proposed. Then, current overseas standards and relevant recent research are reviewed. Finally, it identifies five issues that are crucial to the production of microbiological water standards, and gives the factors that, in this author's opinion, must be addressed. In particular, it is pointed out that the wording of standards must give clear indications about what sampling regime is appropriate, and how compliance is to be judged.

Keywords: Bacteria, microbiology, standard, guideline, criteria, compliance.

DEFINITIONS

The terms "criterion", "standard", and "guideline" do not have common usage in the literature to be referred to: one author's "criterion" is another's "guideline" or "standard". The usage in this report follows that used by a notable worker in this field (Cabelli 1979), and also in Canada (CCREM 1987). It is as follows:

- a "criterion" is the scientific *data* upon which guidelines and standards are based (e.g., a quantifiable relationship between the density of an indicator in a water body and the health risk associated with swimming in that water);
- a "guideline" translates the criteria into a form that *recommends* how a standard should be expressed. This calls for a value-judgement on acceptable health risk;
- a "standard" is what a guideline becomes when it is given *statutory force*. That is, compliance with it is mandated by law.

When referring to the literature it is important to identify the author's usage of these terms. The most common feature is the interchanging of "guideline" and "criteria". For example, the criteria reported by Cabelli (1983) have been used by USEPA (1986) to develop what he calls guidelines, but that agency (because of the wording of the US Clean Water Act) continues to call them "criteria".

INTRODUCTION

Microbiological receiving water standards are a particularly important part of water use planning and in hearing applications for rights to discharge waste into natural waters. They often form the major divisions between the water use classes made available in statute, and can heavily constrain the degree of treatment and location of waste discharges, as in recent water classifications (for Wellington Harbour, Hawke Bay and Poverty Bay). The standards can also influence land developments. There is therefore considerable interest in, and debate on, the use of such standards, e.g., see the discussion of a recent presentation by Professor Loutit (NZWSDA, 1988).

The statutes governing water management in New Zealand have recently been reviewed. One of the issues arising is how best to express microbiological standards for receiving waters, given that the existing standards (in the Water and Soil Conservation Act 1967) are now some years old, and that overseas research has advanced the state of knowledge since then. In May 1989 the author prepared and disseminated a discussion document highlighting the historical development of the present standards - which is poorly documented, and hence poorly understood - and discussing the implications of recent overseas research for any rewriting of standards. This report is based substantially on that document, taking account of comments made on it.

It should be noted that the original discussion document contemplated that any new statutory standards would be written in *explicit* form, as the present standards are. Such standards consist of a numeric limit, often expressed as a percentile, on the results of microbiological enumeration tests for a stated group of organisms carried out over some maximum time period. However in the end-product of the statutes review - the Resource Management Bill introduced to the New Zealand Parliament in late 1989 - the proposed microbiological standards are in *narrative* form. These merely state that certain undesirable outcomes should not occur as a result of microbial contamination. Never-the-less, when developing a "regional plan" (the Bill proposes that these should have statutory force), these standards will generally have to be translated into an explicit form. Thus the original document's discussion on the phrasing and interpretation on explicit standards has been retained. It is hoped that this report may be helpful when the necessary New Zealand microbiological water quality guidelines are prepared: the translation of a narrative standard into a meaningful explicit standard, which will often have to be done when preparing a plan, is not a trivial task.

DEFINING THE PROBLEM

Humans can contract various diseases from microbes in water: from drinking it, swimming in it, or eating shellfish harvested from it. Children seem particularly susceptible. The categories of microbes that can cause disease are well documented elsewhere (e.g., McNeill 1985) and need not be repeated here.

To contain the risk of contracting such diseases, various national and international agencies have developed microbiological guidelines and standards for receiving waters. These are derived from criteria identified in epidemiological studies in which the density of suitable "indicator" organism(s) is correlated with disease risk. An acceptable value of this risk is then selected, by a public authority, and the criteria are used to develop the appropriate guideline or standard. The process the authority uses to decide what risk is acceptable varies: in the United States it involved an iteration with interested parties (as documented by Salas 1986 and CCREM 1987).

Standards should be written and applied in such a way as to be scientifically defensible and enforceable. This means, among other things (e.g., choice of indicator), that the wording of the standard should indicate appropriate sampling regimes, and how compliance is to be assessed. It does need to be borne in mind that compliance with receiving water standards is not an issue that reaches the courts: allegations of offences heard in court concern non-compliance with *effluent* standards. But this does not mean that receiving water standards should be written loosely. After all, the water resource manager's job is to see that the combination of discharges authorised does not cause the receiving water standards to be breached.

It follows that guidelines should give clear guidance on how scientifically defensible and enforceable standards can be written. Included in such a guideline would be discussion on the suitable choice of indicator, and how it is related to disease risk.

There are a number of reviews of potential indicators, e.g., Cabelli 1978, Cooper 1983, Sinton 1988, Sinton *et al.* 1989. Indicators should bear some more-or-less constant relationship to the pathogen level in the water. According to Cabelli (1979) this is valid except where faecal matter is discharged from small populations, or where there is an epidemic in the local contributing population.

The required epidemiological studies, to establish the relationship of indicator density to disease risk, are difficult and expensive. To be really effective they need to be "prospective", in which surveys of water quality are coincident with the use survey: thereafter the users (including non-users, as a control group) have to be followed up to ascertain what illnesses occurred. Few such surveys have been done. It is the nature of such surveys that few, if any, serious (notifiable) illnesses will be found.

Most commonly, epidemiological surveys take the simpler "retrospective" form: they are carried out after the water use occurred and so do not obtain relevant water quality information. Usually, because they often rely on easily available medical records and are not tied to particular locations and periods of use (i.e., water sampling sites and occasions), only notifiable diseases are examined. So, for example, gastrointestinal illnesses would be missed. But because they do obtain data on notifiable diseases one can argue that prospective and retrospective studies are complementary.

Occasionally "predictive" epidemiological models are used. These use available water use and pathogen data to predict disease risk. They tend to be the most speculative type of survey.

Not surprisingly therefore, there has been, and still is, debate over what criteria or standards should be applied.

The issues that need to be addressed in formulating explicit standards are:

1 **Where should standards be applied?**

2 **What indicator species should be used in explicit standards?**

3 **What is the best way to express microbiological standards so as to give clear guidance on the necessary sampling programme and the definition of compliance?**

4 **Should an attempt be made to differentiate human from animal faecal material?** (earlier opinion, e.g., Geldreich 1970, had it that the disease risks of either were more-or-less the same; later opinion says risks from animal faecal material are lower, e.g., Cabelli 1988)

5 **Should there be microbiological standards for the protection of health of aquatic organisms?** (Geldreich *et al.* 1979 were of the view that there should also be criteria in the USA to protect fisheries from pathogens).

HISTORY OF RECEIVING WATER MICROBIOLOGICAL STANDARDS IN NEW ZEALAND

It is appropriate to trace the history of such microbiological standards as New Zealand has had, to clarify the basis for them and document their perceived shortcomings. Parts of this history are blurred because: (i) the organisational changes since the early 1960s have meant that many records cannot be traced, (ii) few publicly available reports appear to have been written, (iii) many of those who were involved have retired and are not available. However invaluable information has come from C.A. Cowie who was closely associated with the drafting of the first standards (those in the 1963 Waters Pollution Regulations) while working in what was then the Ministry of Works (in 1985 he retired from the position of Deputy Director of the Water and Soil Directorate of the now-defunct Ministry). Ian Gunn (University of Auckland) has also provided useful notes. A good general history of the development and application of New Zealand water quality management law, including water standards, has been given by Prendergast (1988).

As will become apparent, New Zealand water microbiological standards are based almost exclusively on USA studies and reports, particularly as described in three reports entitled *Water Quality Criteria* (McKee and Wolf 1963, NTAC 1968, CWQC 1972), and two further reports entitled *Quality Criteria for Water* (USEPA 1976, 1986) (recall that their "criteria" are what others, e.g., CCREM 1987, call "guidelines").

1963 Regulations. The original standards were contained in the Waters Pollution Regulations 1963, issued under the Waters Pollution Act 1953. The Act and the Regulations were administered by the Pollution Advisory Council. The standards in the Regulations were developed by the Council's servicing staff (in what was the Marine Department) and other advisors (notably C.A. Cowie, in what was the Ministry of Works). The classification procedures were explained in a booklet (Pollution Advisory Council 1963). While not stated explicitly in the Regulations or in the booklet, the intention in classifying waters was to promote solutions to existing pollution problems, and also to prevent new problems appearing. Their classifications were therefore implemented only as problems arose or were foreseen in particular areas.

The 19 classifications produced under these Regulations, from 1963 to 1971, are given in Table 1. These were a dramatic step forward in water quality management: for the first time dischargers had to follow statutory procedures to obtain consents, and those consents could place restrictions on their operations. For example, the Bluff-Foveaux Strait classification constrained discharges at the yet-to-be-built Tiwai Point aluminium smelter.

The regulations contained a palette of four classes for inland waters (A, B, C and D), and four for coastal waters (SA, SB, SC and SD). The criterion for class selection, as stated in the Regulations, was the use the Council sought to have promoted or protected. Of the eight, only four (B, C, SA, and SB) contained microbiological standards. The uses they were tied to are expressed as (in s.5):

B	"being water-supply waters in an uncontrolled catchment area";
C	"being waters to which the public have ready access and used regularly for bathing";
SA	"being waters from which edible shellfish are regularly taken for human consumption";
SB	"being waters to which the public have ready access and used regularly for bathing".

Because the areas of such water use are limited, these classifications have only limited areas with coliform standards.

The standard requires that "The coliform bacteria content of the waters shall not consistently exceed (*limit value*) per 100 millilitres". The limit value is 5000, 1000, 50 and 1000 for classes B, C, SA, and SB respectively. The phrase "not consistently exceed" was considered by the Council to be sufficiently flexible so that each case could be treated on its merits; an attempt to be more precise at that stage was seen to be undesirable. It also imposed some uniformity on the expression of the standards, which, as is seen below, was lacking in the USA standards from which they were derived (these used a mixture of medians, means, and sample 90%iles - see Appendix 1 for a definition of sample statistics).

No sampling requirements are indicated, but, as is clear from the wording of the standard, it was intended that the standards be all interpreted as a requirement on percentiles of *time*, not of *samples*. Whether this should be a 50%ile (i.e., median) or a higher percentile wasn't clear: "not consistently exceed" had no further definition. Sampling was intended to be carried out at fixed sites, where the particular uses to be protected were located.

The limit values had their origins as follows.

The B limit value (5000 per 100 ml) was derived from studies by H.W. Streeter for the then United States Public Health Service (USPHS), which related raw water coliform levels to the ability of different treatment processes to remove them (see McKee and Wolf, 1963: 92). The USPHS standard appears to have required the comparison of a mean value with the limit value.

The C limit value (1000 per 100 ml) was based on yet more work by Streeter (1951), on the predicted risk of salmonellosis given such factors as frequency of swimming, the assumption that 10 ml of water will be swallowed by each bather each day, and the probability that this ingestion will cause illness (see McKee and Wolf, 1963: 119). He calculated that the chance of contracting typhoid fever from swimming daily for 90 days in the Ohio River at 1000 coliforms per 100 ml would be 1 in 950, and that of getting diarrhoea-enteritis would be 1 in 50. The many US standards that were based on this work also required the comparison of a mean value with the limit value, and sometimes

an upper sample percentile. There is no suggestion that this standard was based on the USPHS prospective epidemiological studies, which had by then been reported (Stevenson 1953).

The SA limit (50 per 100 ml) was derived from part of the USPHS criteria for "fully approved" shellfish growing waters, requiring that the median coliform Most Probable Number (MPN) not exceed 70 per 100 ml. This limit was reduced to 50 in the Regulations, to include a further margin of safety. The USPHS standard (see McKee and Wolf, 1963: 118) did not indicate a required sampling regime. Whether the median referred to samples or to time is not clear. The rest of the USPHS "fully approved" standard was ignored in class SA (requiring that not more than 10% of samples may exceed 230 per 100 ml, and that the limits need not be applied if the coliforms are not of faecal origin and do not pose a public health hazard). (The "restricted" USPHS standard used 10x the above limits, i.e., 700 and 2300). According to Furfari (1968) the 70 figure arose after studies following a typhoid outbreak in 1924 led to a view that typhoid could be avoided if not more than 50% of the 1 ml portions examined were positive for coliforms. This, using Hoskins equation used to calculate MPN tables, is equivalent to 70 coliforms per 100 ml.

The SB limit value (1000 per 100 ml) appears to have been based on the Ohio River risk calculations of Streeter (1951) supplemented by attainability surveys carried out in 1955 and 1956 on Connecticut coastline beaches and tidal rivers (cited by McKee and Wolf 1963 as Scott 1958 - often cited as 1951, as in Cabelli *et al.*, 1983 and in Salas, 1986). Salas indicates that this limit value may also have been derived from aesthetic considerations: beaches consistently below this limit (more than 80% of the time) remain aesthetically satisfactory, with no signs of sewage pollution. McKee and Wolf (1963: 119) imply that this limit was to be compared with mean sample values.

1971 Amendment Act. The Waters Pollution Act 1953 and its 1963 Regulations were repealed by the Water and Soil Conservation Amendment Act (No. 2) 1971, though existing classifications and their standards remain in force. The Amendment Act now forms part of the Water and Soil Conservation Act. In the Resource Management Bill it is proposed that existing classifications will remain in force (ss.397, 399).

The standards in the 1971 Amendment Act were developed by a committee, some of whose deliberations are reported by Carrie (1973). The standards can be changed by Order-in-Council [s.26C(6)], though they never have been.

Responsibility for classification then first rested briefly (1972-1973) with a transition body, the Water Pollution Control Council (water quantity was the concern of the Water Allocation Council). In 1973 these two were combined into the Water Resources Council, which published a booklet to explain its programme of classification (National Water and Soil Conservation Organisation 1973). It was intended that the whole country should be classified by 1975. As we shall see, court decisions brought this programme to an abrupt halt.[1]

In the Amendment Act, five new coastal classes (SA, SB, SC, SD and SE) and four new inland water classes (A, B, C and D) were introduced. These classes are similar to

---

[1]The Water Resources Council was dissolved (!) in 1984, and the National Water and Soil Conservation Authority took over its responsibilities. These bodies were all serviced by the Ministry of Works and Development. The Authority and the Ministry were dissolved altogether in 1988, leaving catchment authorities with the responsibility of water classification. In 1989 the catchment authorities were dissolved, and their functions taken over by 13 regional government agencies.

those in the Regulations (the main differences being that explicit temperature change and descriptive colour and clarity standards were introduced and some numerical limits were changed slightly). Accordingly, the same class labels were used, and one new class (SE) was added (for ocean outfalls discharging comminuted waste). The symbol "X" could be attached to any class, to signify that the waters are sensitive to enrichment.

But in the Amendment Act the classes were no longer tied to water uses: that this was a fundamental change did not become apparent for a year or two. A High Court (Administration Division) decision by Cooke J. [1976 1 NZLR 1] stated that water use was not the sole criterion for selection of a water class in classifying water. In effect, the highest possible class was to be used. That the classes were designed to protect certain water uses was ignored. As a result of this decision the Water Resources Council felt that a large number of classifications and reclassifications in various stages of production would not stand on appeal. They were withdrawn.

The 11 classifications carried out under this Act, some before and some after the Cooke decision, are given in Table 2, which also gives details of those that were withdrawn. Those done after the Cooke decision have been for coastal waters only. They have large areas of water with faecal coliform or total coliform standards, raising large questions about the definition of compliance with microbiological standards (discussed later).

Pressure for the incorporation of the 1963 Regulations into the Water and Soil Conservation Act included the need to provide for prosecutions, and the desire by consulting engineers and local authorities to give a more precise definition of microbiological standards than the phrase in the Regulations: "... shall not consistently exceed ...". A need was seen also to take into account more recent research and so revise the type of coliforms measured and also the limits applied to them. Accordingly, the coliform type and limit values applying to all the new classes, which are those applying to classifications issued since 1972, were changed to:

Class	*Type*	*Limit value*
B	Total and faecal	10000 and 2000
C	Faecal	200
SA	Total	70
SB	Faecal	200

These classes all require that: "Based on not fewer than 5 samples taken over not more than a 30-day period, the median value of the (*type*) coliform bacteria content of the waters shall not exceed (*limit value*) per 100 millilitres". For brevity this will be referred to as the "5-in-30 rule". Once again, this requirement imposed a uniformity that was lacking in the studies from which they were derived (these used "log means", arithmetic means, time medians, and upper sample percentiles - and note that "log mean" is used incorrectly, see Appendix 1). The attraction of the "5-in-30" rule to the committee that recommended it was that it offered the prospect of being able to track the impact of a treatment plant on downstream water quality over some reasonably long compliance assessment period.

Again, no sampling requirements were indicated, but a careful reading shows that the rule is a requirement on percentiles (medians) of time, not of samples: the limits are for the median bacterial content "of the waters", not of samples. And the intention was to sample at fixed sites. The limit values had their origins in the second *Water Quality Criteria* report (NTAC 1968 - see Carrie 1973 - whose standards were designed to protect certain water uses). This report was actually compiled by five subcommittees,

for five categories of water use. Not surprisingly, their criteria were expressed in different ways. Details are as follows.

The water supply standard (equivalent to class B) was based on the expectation that a defined water treatment plant could process such water to meet then-current US drinking water standards (NTAC 1968: 20-21). Advances in water treatment technology led to the old coliform standard being doubled. The limit values are NTAC's "permissible criteria". The NTAC standard was based on "... monthly arithmetic averages based on an adequate number of samples": it did not require the "5-in-30" rule.

The primary contact recreation (bathing) criteria (equivalent to classes SB and C) were based on the epidemiological studies of the USPHS from 1948 to 1950, as reported by Stevenson (1953). These were paired beach studies (one beach being more polluted than the other): surveys included a non-swimming control group. The sites were at: Lake Michigan at Chicago (5124 persons); the Ohio river at Dayton Kentucky, with a less polluted "beach" being a nearby swimming pool (7520 persons); and on Long Island Sound, New York, at the New Rochelle and Mamaroneck beaches (9520 persons). No correlation between illness rate and bacterial quality was found at the marine beaches, but two were found at the freshwater sites. (i) At the more polluted Chicago beach the illness rate from swimming in a 3-day period when the average coliform content was 2300 MPN/100 ml was "significantly" higher than for three other days at 43 MPN/100 ml. The significance level attained was $p = 0.01$. Note that this "average" coliform content (at p. 537) is described on p. 535 as a "logarithmic average" (presumably, the geometric mean), which is unfortunate: for skewed data the two statistics would be quite different. (ii) At the Ohio River beach with a median coliform content of 2700 MPN/100 ml over the whole study period, the number of swimmers with gastro-intestinal illness was significantly higher than expected ($p = 0.05$).

The NTAC 1968 report (at p. 12) used these findings to develop their primary contact criterion, saying that they "showed an epidemiologically detectable health effect at levels of 2300-2400 coliforms per 100 ml." The 2400 should presumably have been 2700. Wishing to have the standard expressed in terms of faecal coliforms, they noted that later Ohio River studies (unspecified) showed a faecal:total coliform ratio of 18%, and applying a safety factor of two the faecal coliform limit was set at 200 per 100 ml. (From CWQC 1973: 31 it appears that the 18% figure was reported by Geldreich 1966). The NTAC criterion explicitly allowed MPN or MF enumeration and introduced the "5-in-30" rule. Indeed it was the only subcommittee to recommend this rule. The limit was expressed as a "log mean" (which is actually a geometric mean: again, see Appendix 1), not as a median, and not more than 10% of samples were to exceed 400 faecal coliforms per 100 ml. No justification for the "5-in-30" rule or the 10% of samples requirement was given.

The shellfish standard (equivalent to class SA) was based on the then-current US National Shellfish Sanitation Program's Manual of Operation (NTAC 1968: 37). It required the coliform median MPN to not exceed 70 per 100 ml, and not more than 10% of samples were to exceed an upper figure of 230 or 330 depending on whether the 3- or 5-tube test was used. Sampling was to be "... in those portions of the area most probably exposed to fecal contamination during the most unfavourable hydrographic and pollution conditions." Other than that, no sampling regime was prescribed.

Draft consolidation of water law. In recent years there have been attempts made to redefine standards. This has been guided in part to overcome the impasse that resulted from the decision of Cooke J. Associated with this of course was perceived shortcomings of the microbiological standards. These had particularly to do with the inappropriateness of total coliforms, the possibility of using microbes other than faecal

coliforms as indicators, and lack of guidance in how to interpret compliance with the standards as written.

The first major step was the report of the Water Quality Criteria Working Party (WQCWP 1981). They proposed five freshwater and five coastal water classes being tied to particular water uses. Most importantly, the process of classification was seen to be optional and subordinate to "water management statements", in which community aspirations for water use would be identified. The proposed classes were given labels different from those in the Act. These proposed standards (with a few changes) were included in the draft Water and Soil Conservation Bill of June, 1986, prepared by the Ministry of Works and Development. That Bill was never introduced to Parliament, but has served as a building block for the Resource Management Bill.

Classes in the draft Water and Soil Conservation Bill containing microbiological limits are:

R	"being water for regular public bathing";
W	"being water for a source for public water supply or for the preparation and processing of food for sale for human consumption where treatment at least equivalent to flocculation, filtration, and disinfection could be reasonably expected";
CR	"Being coastal water for regular public bathing purposes";
CS	"Being coastal water from which edible shellfish are regularly taken for human consumption or waters in which shellfish are cultivated or farmed".

These proposed standards were all based on faecal coliforms.

Classes R, W and CR require that "The median faecal coliform bacteria concentration shall not exceed (*limit value*) per 100 millilitres based on a minimum of one water sample taken on each of five separate days over not more than a 30 day period, nor shall more than 10% of the samples taken on separate days during any 30 day period exceed (2 x *limit value*) per 100 millilitres." The limit values are 200, 200 and 2000 for classes CR, R and W respectively [note that in the draft Bill the (2 x limit value) for class W is given incorrectly as 400; it should be 4000].

For class CS the requirement is somewhat different: "The median faecal coliform bacteria concentration shall not exceed 14 MPN (Most Probable Number) per 100m (sic) millilitres based on a minimum of one water sample taken on each of 10 consecutive days when the risk of contamination is greatest, and not more than 10% of the sample (sic) shall exceed 43 MPN per 100 millilitres."

A notable omission from these statements is the explicit requirement that the limits be on the bacterial content *of the waters*.

With two exceptions the proposed bathing water standards (R and CR) are the same form as those in NTAC (1968), and repeated in the 1976 *Quality Criteria for Water* report (USEPA 1976). The exceptions are that the median has been used in place of the "log mean", and sampling is to be on separate days. It is to be noted that by this time the USPHS prospective epidemiological studies reported by Stevenson (1953) and their interpretation by NTAC (1968) were being criticised on a number of counts (e.g., CWQC 1973, Cabelli *et al.* 1975, Moore 1975). In fact the USPHS work was largely abandoned in USEPA (1976) as a rationale for the NTAC criterion. Instead, appeal was made to a relationship between faecal coliforms to the frequency of *Salmonella* isolations in surface waters [the 1973 CWQC report had appealed to this relationship also, citing Geldreich 1970 (not cited by USEPA!), but did not feel it strong enough to recommend a criterion].

A letter from E.E. Geldreich to D. Till in Appendix 1 of WQCWP (1981) sheds some light on the origin of the limit on 10% of samples requirement in the US bathing criteria from NTAC 1968 onward. A member of the NTAC bathing water subcommittee (Lee McCabe) stated that it resulted from his (unpublished) analysis of bathing beach and epidemiological data of increased illness among bathers, concluding that "... the risk was significant when more than 5% of the samples exceeded the 400 faecal coliforms per 100 ml value." Concern about practicality of sampling a beach 20 times per month convinced the subcommittee to reduce 5% to 10%. (Given the importance of this observation, it is a great pity that this analysis has not been published). Clearly, the log mean value was intended for monitoring long-term conditions, while the 10% of samples statistic was for immediate questions about whether to open or close a beach.

The proposed water supply standard (W) was based on the NTAC (1968) criterion, but with the median replacing the arithmetic average, and insertion of the "5-in-30" rule.

The proposed shellfish harvesting waters standard (CS) was based on that in USEPA (1976) which was based on the previous median standard of 70 total coliforms per 100 ml. A study by the National Shellfish Sanitation Program collected coliform data from 15 States and two Canadian Provinces. From about 3500 samples USEPA (1976) reports that "... 70 coliform MPN per 100 ml at the 50th percentile was equivalent to a fecal coliform MPN of 14 per 100 ml. The data, therefore, indicate that a median value for a faecal coliform standard is 14 and the 90th percentile should not exceed 43 for a 5-tube, 3-dilution method ...". The text (at p. 48) says that, where possible, samples should be collected when pollution could be expected to be maximum. The reason for including in class CS the requirement that samples be taken on "each of 10 consecutive days" is not given: it is not in USEPA (1976), was not recommended by WQCWP (1981), and appears not to appear in overseas food standards with which New Zealand exports have to comply.

CURRENT OVERSEAS GUIDELINES AND STANDARDS

Salas (1986) and McNeill (1985) have presented summaries of guidelines and standards used in the US States, Canada, Europe and some other countries. This covers water supply waters, bathing waters, shellfish waters, and others besides (swimming pool waters and "protection of indigenous organisms"). Sample percentiles, not time percentiles are used commonly. It appears that New Zealand has not been alone in being heavily influenced by the NTAC (1968) guidelines. It is notable that, except for Europe, the "5-in-30" rule has been adopted in one form or another.

The EEC appears not to have shellfish harvesting water standards (though it does have them for shellfish flesh - Council Directive 79/923/EEC). Its bathing water standards (Council Directive 76/160/EEC, Pedini 1976) place a limit on 80% or 95% of fortnightly samples taken over the bathing season. They include limits on fecal streptococci (100 per 100 ml), *Salmonella* (0 per litre), and enterovirus (0 PFU per litre).

Since the Salas and McNeill reviews appeared there is now a likelihood that the EEC Directive will be changed (Jones and Kay 1989), from 80% or 95% below 2000 FC/100 ml to 95% below 1000 *E. coli*/100 ml. This would probably be a more restrictive standard.

In 1986 the USA guidelines (= their "criteria") were reviewed (USEPA 1986a).

The water supply water guideline was unchanged from that in CWQC (1973). The shellfish harvesting water guideline also remained unchanged, except that, inexplicably, the requirement to sample at times when pollution is expected to be worst was omitted.

An oversight, surely: D. Till (Department of Health, *pers. comm.*) states that the international health standards do so require.

The USA bathing water guidelines have changed substantially, being derived from work on marine waters by Cabelli (1983) and on freshwaters by Dufour (1984). In particular, faecal coliforms have been replaced by *E. coli* and enterococci, enumerated with specially developed membrane filtration techniques (USEPA 1985). These guidelines call for "... a statistically sufficient number of samples (generally not less than 5 samples equally spaced over a 30-day period) the geometric mean of ... should not exceed ...". For freshwater the limits are either 126 *E. coli* per 100 ml, or 33 enterococci per 100 ml. For marine waters the limit is 35 enterococci per 100 ml. Single sample maximum allowable densities are also given for various intensities of swimming use.

These guidelines were developed from linear regression equations of the logarithm of mean densities on illness rate. "Acceptable Swimming Associated Gastroenteritis Rates per 1000 swimmers" were adopted: of 8 for freshwater and 19 for marine water. These risks were adopted as they were calculated to be the same as the risks implied by the previous faecal coliform guideline (USEPA 1986a). In a draft of the guideline these risks were initially at a lower level, but they were raised after submissions were received (CCREM 1987, pp.2-3, 2-4).

No justification was given for the "statistically sufficient" claim, nor for the inclusion of the suggestion that samples should be "equally spaced" in time over a 30 day period. This latter requirement is of course an embellishment of the "5-in-30" rule introduced by the NTAC (1968). Sampling was suggested to be in dry weather and be weekly, bi-weekly, or monthly according to intensity of use (USEPA 1986b). This is a little difficult to reconcile with the "5-in-30" rule!

RECENT RESEARCH

I am not aware of relevant recent published research that would affect standards for water supply waters. There are varying findings about the relative level of indicators and pathogens in shellfish harvesting waters and in shellfish flesh (e.g., Power et al. 1988). In USEPA 1986a it was noted the the potential for shellfish harvesting waters criteria being base on enterococci and *E. coli* was to be examined. Research does not yet seem to be pointing to improved standards for shellfish harvesting waters.

The active research that I am aware of has been for bathing waters, initiated by Cabelli and his co-workers (especially Dufour), for the USEPA. Cabelli's work was for marine waters, Dufour's was for freshwaters. Their work is summarised well in USEPA (1986b). In brief, prospective epidemiological studies of swimmers and non-swimmers for a number of paired USA beaches and freshwater sites were conducted. In each case one of the pair of sites was close to point sources of treated effluents, most of which were disinfected. The other site, further away from point sources, as well as a non-swimming cohort, were used as controls. The survey design attempted to avoid the pitfalls of the USPHS studies reported by Stevenson (1953), as explained in Cabelli *et al.* (1975). Some 26700 respondents were used in the marine studies, and 45500 in the freshwater studies.

The early marine studies, on two New York city beaches, measured a number of indicators to determine which ones correlated best with Swimming-Associated Gastroenteritis Rate. This work indicated that enterococci and *E. coli* were correlated best and so only these two, and faecal coliforms (because of its historical use), were used in the subsequent marine and freshwater beach studies.

The fundamental result obtained was that a significant Swimming-Associated Gastroenteritis Rate was always observed at the more polluted beaches, but not at the less polluted beaches. So it was concluded that "*... there is a measurable and significant risk of acute gastroenteritis associated with swimming in marine waters which are contaminated with human faecal wastes to levels less than those which would be aesthetically unacceptable.*" (Cabelli 1988). Skin, ear, eye and respiratory symptoms were speculated to be largely attributable to contamination between bathers in conditions of heavy usage and poor water exchange. These findings replace those reported by Stevenson (1953): in particular, they show that there *is* a risk of illness at marine beaches: when indexed with enterococci, the risk is higher for marine waters.

Since the EPA studies were done, further prospective epidemiological studies have been carried out on marine beaches in Egypt, England, France, Israel, Spain and Hong Kong (Cabelli 1988, Jones and Kay 1989). Further freshwater studies have been carried out in Canada and Connecticut (Cabelli 1988). The "Cabelli protocol" (i.e., his prospective epidemiological study design) is being tested at New Jersey beaches now (Jones and Kay, 1989).

These latter authors make four criticisms of this protocol, pointing out that apparently different results have been obtained in other studies, particularly the freshwater studies in Lake Ontario, Canada. Without going into too much detail, I think these criticisms are rather shallow, and had already been addressed in Cabelli (1988). There are four points to be made:

- Jones and Kay (1989) claim that the Canadian work (Seyfried *et al.* 1985) found that "total staphylococci proved the best indicator", and that this is at variance with Cabelli and Dufour who found that *E. coli* or enterococci correlate best with illness rate. But Seyfried *et al.* did not measure either *E. coli* or enterococci in their studies; conversely, Dufour and Cabelli did not measure staphylococci. So a finding that staphylococcal infections correlate highest (at heavily used, poorly flushed beaches where staphylococcal cross-infections can be expected) says nothing about faecal contamination;
- Cabelli and Dufour do not claim that their results apply world-wide: indeed, Cabelli's work in Egypt was not used to develop the criteria in USEPA (1986a), because of an expectation that immunity rates in Egypt would be markedly different from those in the USA (but note that New Zealand immunity rates could be expected to be more similar to the USA rates than would Egyptian rates);
- Dufour's (and Cabelli's) beaches were often impacted by disinfected effluents: the Canadian studies appear not to have been;
- Dufour and Cabelli collected large statistical samples: much larger than all the other studies except that in Spain (which is flawed anyway - see Jones and Kay 1989). With larger samples more correlations will appear to be "significant" (see Appendix 2).

DEVELOPING GUIDELINES FOR NEW ZEALAND

We can now turn to the five issues identified at the start. The following discussion does not attempt to resolve them, but does indicate what must be taken into account in attempting to do so.

1. Where should standards be applied?

All water quality guidelines and standards are derived to protect and promote desirable uses of water. Ideally, the "desirable" uses are identified through some consultative planning process. It is therefore clear that in a plan, bathing beaches and shellfish harvesting waters should have appropriate microbiological standards attached. The recent proliferation of windsurfing means that consideration has to be given to applying

a bathing water standard over considerably larger areas than before. It has been the case that water supply waters have had standards also (except for protected water supply catchments), but WHO do not now have a guideline for these waters - it is believed that suitable treatment of a raw water can bring it up to the required drinking water standard.

Classifications prior to the Cooke decision left the waters not having these three uses without microbiological standards. This is generally the bulk of the water body. Subsequent classifications, which have all been for coastal waters (e.g., Hawke Bay), have microbiological standards over practically all water, because his decision requires standards to be set as "high" as is practicable. As a result, only areas very close to major outfalls, some estuaries and lagoons, and wildfowl areas are exempt. While many are pleased to see a high standard (SA) applied over large areas of marine water, it should be noted that this standard was derived for shellfish harvesting waters, and that certainly does not occur in many areas currently being given that classification. It would appear that the main *technical* reason for imposing an explicit microbiological standard over such waters would be for the protection of indigenous species (as claimed by Geldreich *et al.* 1979).

There are difficulties applying microbiological standards to many freshwaters: the nature of the land-use means that bathing water standards would be breached (though they would not if the indicator were specific to human faecal matter).

2. What indicator species should be used in explicit standards?
If any guideline is made for water supply waters (other than protected water supply catchments, from which the public is excluded), faecal coliforms is the appropriate choice. For shellfish harvesting waters there appears to be little option but to stick with faecal coliforms. The methods used have to be MPN, or a membrane filtration technique that returns similar results. For bathing water standards, Cabelli and Dufour's work is a much more solid foundation than is the old USPHS study upon which the present faecal coliform standard is based. If their membrane filtration technique (USEPA 1985) is feasible here, and if it really does produce good results for turbid samples, we have to seriously consider changing to *E. coli* and enterococci. Note that the Canadians (CCREM 1987) have not yet changed from faecal coliforms. They will not do so until standard laboratory protocols are established. With the demise of the New Zealand National Water and Soil Conservation Authority no agency now has such responsibility. It has been well argued (Wood 1988) that such a protocol is necessary in New Zealand.

Whatever methods are used, they have to give similar results to those used in the epidemiological studies upon which the standards are based. *Methods that recover many less, or many more, are inappropriate.*

There does not seem to be sufficient evidence to support particular viral standards, though it would be wrong to ignore viruses: that they cause disease and can be very persistent (more so than bacteria) is well-known. Some of the difficulties are that their enumeration is not a simple task, and most often they are absent, only appearing when an outbreak occurs in the contributing population. Faecal bacteria (e.g., *E. coli*) at least are reliable indicators of the presence of sewage. Guidelines can consider leaving in some narrative statement about viral contamination, leaving each case to be treated on its merits. There is yet much profitable research to be done on this topic (e.g., Cabelli 1988 speculated that the viral agents of acute gastroenteritis may be ubiquitous and relatively constant). It is entirely possible that it will become feasible to set and monitor such standards routinely.

3 What is the best way to express microbiological standards so as to give clear guidance on the necessary sampling programme and the definition of compliance?

When we consider compliance, there's more than just some limit to consider: the length of the compliance period and the compliance rule (e.g., median not to exceed ...,"5-in-30" rule, upper limits) must be considered also. I think that we ought to continue to keep the statistical expression of the standard uniform across the classes and to make it crystal clear that we're talking of requirements on percentiles of *time*, not percentiles of *samples*. This should be the case for upper percentiles too - even though everybody else seems to use upper percentiles of samples.

The problem about limits on sample percentiles is that such a requirement says *nothing* about the required sampling. But if one makes a requirement on percentiles of time, then sampling has to be directed toward testing the hypothesis that the limit was not exceeded for more than the required period of time. That is, *sampling has to be random.* Actually, this is implied in the wording of the B, C, SA, and SB standards in the Water and Soil Conservation Act: the requirement is on the median bacterial content "of the waters".

But random sampling can be silly: it would mean for example that a lot of our bathing waters sampling would be at night! The problem here of course is that in formulating the standard, the way in which data were gathered in the USPHS prospective epidemiological studies upon which it was based has been ignored: during the day, when people were swimming. So why not require that stratification of sampling in the expression of the standard? For example, sampling could be directed toward bathing periods (as in the USA enterococci standard), or toward periods when contamination of shellfish harvesting waters may be expected to be at a maximum.

And we don't need to stick to 30 days, or (a minimum of) 5 samples. For bathing waters the EEC approach of sampling during the bathing season enables one to have much smaller risks of reaching false conclusions about compliance, because there are many more than 5 samples. With only 5 samples, there is poor confidence in assessing compliance (see Appendix 3). There has been some comment that these considerations (as outlined in Appendix 3) are just statistical niceties, bearing little consequence to practical realities. Not so! If one is taking small samples from a vast environment, and making inferences about compliance, *the risks of reaching false conclusions cannot be avoided.* Proper use of statistics, both in designing standards and in sampling to assess compliance can at least quantify these risks and enable them to be minimised in some way. This is already used in ocean outfall compliance testing in New South Wales (Kaye and Webb 1988): they also note that only 5 samples carries large risks of reaching false conclusions. Guidelines should give simple procedures for balancing error risks.

A further question is whether standards applied to coastal waters should state whether fixed site sampling is to be used. This question has arisen in the recent Hawke Bay and Poverty Bay classifications, where class SA and SB waters are in close proximity to waste outfalls. A waste plume wafts around in the ocean, so if one samples always in the middle of the plume (moving-site sampling) the chance of finding a breach of a standard is very much higher than if sampling is at fixed points. Note that a recent Planning Tribunal decision (on appeals against the Hawke Bay classification by the Hawke's Bay Regional Water Board, Decision No. W28/89 of 6 April 1989) has indicated that under the Water and Soil Conservation Act sampling should indeed be at fixed sites.

And finally, the upper percentile is often based on twice the central limit value (e.g., 200 and 400 FC/100 ml in the R, W and CR standards in the 1986 Draft Bill). This doubling is based on the unpublished work of Mr McCabe (a member of the bathing water criteria subcommittee of NTAC 1968). But published work shows that data show a much greater ratio of upper to middle percentiles. Pike and Gameson (1970) have pointed out that the USPHS data can exhibit a 9:1 ratio of 90%ile to 50%ile. Withers (1980) reported a 80%ile:50%ile ratio for Wellington Harbour FC of 4:1.

River FC data that I have for the Waikato River suggest a 90%ile:50%ile ratio of about 5:1.

Perhaps to avoid this problem, the Canadian Guideline requires that if any single sample exceeds 400 FC per 100 ml then "resampling should be performed" (CCREM 1987). What should be done then is not clear.

If time medians are used in standards, it would seem that an upper time percentile (e.g., a maximum) is needed also. The question is not so clear if standards are expressed as a requirement on the geometric mean: a geometric mean after all does take some account of extreme values: a median does not. Note that if any sample value is zero, the sample geometric mean is then zero also, regardless of the magnitude of the other sample results. Guidance is also needed on how to handle and interpret "less than" and "greater than" data so common to bacteria enumerations. Most existing software cannot easily cope with such censored data, and the < or > sign quickly becomes detached from the record, which is then corrupted, at times alarmingly (a recent VAX-based water quality data storage system - AQUAL, McBride 1989 - stores such data as an entire unit, so that detachment cannot occur).

4. Should an attempt be made to differentiate human from animal faecal material? According to Noonan *et al.* (1988) we don't really know too much yet about the public health significance of meatworks wastes, though the weight of opinion is that human faecal matter carries quite a higher risk.

5. Should there be microbiological standards for the protection of health of aquatic organisms? This must await further research and input.

REFERENCES

Cabelli, V.J., 1978. New standards for enteric bacteria, in *Water Pollution Microbiology*, **2**, R. Mitchell (ed.), Wiley, New York: 233-271.

Cabelli, V.J., 1979. Evaluation of recreational water quality, the EPA approach, in *Biological Indicators of Water Quality*, A. James and L. Evison eds., Wiley, Chichester: 14-1 - 14-23.

Cabelli, V.J., 1983. *Health effects criteria for marine recreational waters*. U.S. Environmental Protection Agency, Report **EPA-600/1-80-03**, Cincinnati, OH, 98 p.

Cabelli, V.J. 1988. Swimming-associated illness and recreational water quality criteria. *Proceedings of the conference on water quality and management for recreation and tourism*, IAWPRC/AWWA, Brisbane, Australia, July 10-15 (unpaginated).

Cabelli, V.J., Levin, M.A., Dufour, A.P., and McCabe, L.J., 1975. The development of criteria for recreational waters, in *International Symposium on Discharge of Sewage from Sea Outfalls*. A.L.H. Gameson (ed.), Permagon Press, London: 63-73.

Cabelli, V.J., Dufour, A.P., McCabe, L.J., Levin, M.A.1983. A marine recreational water quality criterion consistent with indicator concepts and risk analysis. *Journal of the Water Pollution Control Federation* **55(10)**: 1306-1314.

CCREM, 1987. *Canadian water quality guidelines*. Canadian Council of Resource and Environmental Ministers, Environment Canada.

Carrie, M.S., 1973. Coliforms and water quality legislation. *Soil and Water* **10(2)**, Published by the Ministry of Works and Development for the National Water and Soil Conservation Organisation: 20-23.

Cooper, A.B., 1983. The uses of microbiological water quality parameters, in *Design of Water Quality Surveys*, R Hoare (ed.). Water and Soil Miscellaneous Publication **63**, Ministry of Works and Development, Wellington: 63-73.

CWQC, 1973. *Water quality criteria*. National Academy of Sciences National Academy of Engineering, Report of the Committee on Water Quality Criteria, U.S. Environmental Protection Agency **EPA-R3-73-033**, Washington, D.C., 594 p.

Dufour, A.P., 1984. *Health effects criteria for fresh recreational waters*. U.S. Environmental Protection Agency, Report **EPA-600/1-84-04**, Cincinnati, OH, 33 p.

Furfari, S.A., 1968. History of the 70 MPN/100ml standard. Memo, Northeast Marine Health Sciences Laboratory, USA.

Geldreich, E.E., 1966. *Sanitary significance of fecal coliforms in the environment*. Federal Water Pollution Control Administration Report **WP-20-3**, Washington D.C., 122 p. (not seen).

Geldreich, E.E., 1970. Applying bacteriological parameters to recreational water quality. *Journal of the American Water Works Association* **62(2)**: 113-120.

Geldreich, E.E, Clem, J.D., Goldmintz, D., Graikoski, J.T., Schillinger, J.E., Stuart, D.G., and Swanson, D.L., 1979. Pages 82 - 89 in *A review of the EPA Red Book Quality criteria for water*. Thurston, R.V., Russo, R.C., Fetterolf, C.M.Jnr.,Edsall, T.A., and Barber, Y.M., Jnr (eds.), 1979. Water Quality Section, American Fisheries Society, Bethesda, MD.

Jones, F., Kay, D., 1989. Bathing waters and health studies. *Water Services* **93**: 87-89.

Kaye, R.B. and Webb, A.T., 1988. Workable compliance testing for ocean outfall performance, in, Proceedings of the *International Conference on Marine Disposal of Wastewater*, NZWPRC/Wellington Regional Council (unpaginated).

McBride, G.B., 1987. Sock'em with OCCAM. *Water Quality Centre Internal Report* **87/09**, 12 p.

McBride, G.B., 1989. AQUAL user's and reference manual, version 1.0. Division of Water Sciences, Department of Scientific and Industrial Research, New Zealand.

McBride, G.B., and Pridmore, R.D., 1987. A statistically defensible procedure for expressing percentile effluent standards, in Proceedings of the New Zealand Water

Supply and Disposal Association Annual Conference, Rotorua, Published by NZWSDA: 242-257.

McNeill, A.R., 1985. Microbiological water quality criteria: a review for Australia. *Technical Paper* **85**, Australian Water Resources Council, 566 p.

McKee, J.E., and Wolf, H.W., 1963. *Water quality criteria, 2nd. ed.*, State Water Quality Control Board Publication **3-A**, 538 p.

Moore, B., 1975. The case against microbial standards for bathing beaches, in *International Symposium on Discharge of Sewage from Sea Outfalls*. A.L.H. Gameson (ed.), Permagon Press, London: 103-109.

NTAC, 1968. *Water quality criteria.* Report of the National Technical Advisory Committee to the Federal Water Pollution Control Administration, 234 p.

National Water and Soil Conservation Organisation, 1973. *What you should know about water use.* WASCO **17**, published by the Ministry of Works, Wellington.

Noonan, M.J., Crichton, C.M., and Taylor, K.J.W., 1988. What is the significance of an indicator bacterium near a marine outfall?, in, Proceedings of the *International Conference on Marine Disposal of Wastewater*, NZWPRC/Wellington Regional Council (unpaginated).

NZWSDA, 1988. Relevance and application of faecal coliform tests - a discussion at the NZWSDA Conference, Dunedin, 18 August, 1988. New Zealand Water Supply and Disposal Association Newsletter **53**: 21-27.

Pedini, M., 1976. Council Directive of 8 December 1975 concerning the quality of bathing water (**76/160/EEC**). *Official Journal of the European Communities*, No **L 31**, 7 p.

Pike, E.B., and Gameson, M.A., 1970. Effects of marine sewage disposal. *Water Pollution Control*, **69**: 355-382.

Pollution Advisory Council, 1963. *The Waters Pollution Regulations 1963 explanatory booklet.* Pollution Advisory Council Publication **9**. Published by the Marine Department, 30 p.

Power, F.M., Wilson, A.E., and Bedford, G., 1988. Chlorination of a short sea outfall secondary-treated effluent. Effects on faecal coliform numbers in the coastal environment, in, Proceedings of the *International Conference on Marine Disposal of Wastewater*, NZWPRC/Wellington Regional Council (unpaginated).

Prendergast, P.F., 1988. Policies and legislation governing effluent discharges in New Zealand, in, Proceedings of the *International Conference on Marine Disposal of Wastewater*, NZWPRC/Wellington Regional Council (unpaginated).

Salas, H.J., 1986. History and application of microbiological water quality standards in the marine environment. *Water Science and Technology,* **18**: 47-57.

Scott, W.J., 1958. Sanitary study of shore bathing waters.*Bull. Hyg.*, **33**: 353. (not seen).

Seyfried, P.L., Tobin, R.S., Brown, N.E., and Ness, P.F., 1985. A prospective study of swimming-related illness. II. Morbidity and the microbiological quality of water. *American Journal of Public* Health, **75(9)**: 1071-1076.

Sinton, L.W., 1988. Bacteriophages as indicators: a preliminary review and comments on research needs. Contract Report No. **82**, Geophysics Division, Department of Scientific and Industrial Research, Christchurch, New Zealand.

Sinton, L.W., Hastie, C.M., and Cruikshank, F.T., 1989. Faecal streptococci as faecal indicators: a review. Contract Report No. **125**, Geophysics Division, Department of Scientific and Industrial Research, Christchurch, New Zealand.

Stevenson, A.H., 1953. Studies of bathing water quality and health. *American Journal of Public Health*, **43**: 529-538.

Streeter, H.W., 1951. Bacterial quality objectives for the Ohio River: a guide for the evaluation of sanitary conditions of waters used for potable supplies and recreational uses. *Ohio River Valley Water Sanitation Commission*, Cincinnati (not seen).

USEPA, 1976. *Quality criteria for water.* United States Environmental Protection Agency, Washington D.C.

USEPA, 1985. Test methods for *Escherichia coli* and Enterococci in Water by the Membrane Filtration Procedure. Report **EPA 600/4-85-076**, United States Environmental Protection Agency, Cincinnati, OH, 25 p.

USEPA, 1986a. Quality criteria for water 1986. *Report* **EPA 440/5-86-001**, United States Environmental Protection Agency, Washington D.C. May 1.

USEPA, 1986b. Ambient water quality criteria for bacteria - 1986. Report **EPA 440/5-86-002**, United States Environmental Protection Agency, Washington D.C.

Williams, D.A.R., 1980. *Environmental Law*. Butterworths, Wellington: 117-118.

Withers, C.S., 1980. Faecal coliform standards: some comments for the Water Quality Criteria Working Party meeting of 2.7.80. Department of Scientific and Industrial Research, Applied Mathematics Division, Victoria University, Wellington, 12 p.

WQCWP , 1971. The report of the Water Quality Criteria Working Party. *Water and Soil Miscellaneous Publication* **25**, Ministry of Works and Development, Wellington, 31 p.

Wood, S., 1988. Review of microbiological indicator systems and criteria for fresh and marine recreational waters, in Proceedings of *Water Conference 1988*, sponsored by Institute of Professional Engineers of N.Z./Royal Society of N.Z., Dunedin, 15-19 August: 281-290.

Zar, J.H., 1984. *Biostatistical Analysis*. 2nd. ed., Prentice-Hall, 718 p.

TABLE 1 CLASSIFICATIONS PRODUCED BY POLLUTION ADVISORY COUNCIL UNDER THE WATERS POLLUTION REGULATIONS 1963[a]

1	Whangarei Harbour
2	Waikato River and catchment
3	Kaituna - Rotorua
4	Tarawera River (including Lake Tarawera)
5	Tauranga Harbour[b]
6	Ohope - Ohiwa Harbour
7	Napier - Hastings
8	Kaupokonui River
9	Lower Wanganui River
10	Lower Manawatu River
11	Waikanae River
12	Titahi Bay
13	Hutt River
14	Ruamahanga River
15	Nelson - Waimea Inlet
16	Opawa River
17	Lower Waimakariri River
18	Mataura River
19	Bluff - Foveaux Strait

[a] Standards applying to these waters are those contained in the Regulations, **not** those contained in the Water and Soil Conservation Act [see *Canterbury Frozen Meat Co Ltd v North Canterbury Catchment Board* (1977) 6 NZTPA 280, 287 - as cited by Williams 1980].

[b] The Tauranga Harbour classification was revoked in 1974, in expectation of a final reclassification (which was subject to appeal). In the interim, following Cooke J.'s 1975 decision, the reclassification was withdrawn, along with other classifications in various stages of finality. So the harbour has not been classified since 1974.

Source: Evidence presented by W.R. Howie to Cooke, J., and P. Prendergast (*pers. comm.*)

TABLE 2 CLASSIFICATIONS PRODUCED UNDER THE WATER AND SOIL CONSERVATION ACT 1967, THROUGH THE PROVISIONS BROUGHT INTO THE ACT BY THE AMENDMENT ACT (NO. 2) 1971[a]

1	Bay of Islands[b]
2	Lake Rotorua[c]
3	Hawke Bay[d]
4	Poverty Bay and coastal waters[d]
5	Lake Horowhenua[e]
6	Manawatu River (whole catchment)[f]
7	Wellington Harbour[d]
8	Timaru coastal waters[d]
9	Southland[b]
10	Bluff - Foveaux Strait[e]
11	Stewart Island and coastal waters[d]

[a] Classifications were produced by the Water Resources Council or the National Water and Soil Conservation Authority, except 3,4 and 6 (the Act was changed to allow Regional Water Boards to classify water after 1/4/88).
[b] Parts of these were the subject of Cooke's decision. After the decision was released the following were withdrawn by the Water Resources Council: existing preliminary classifications (Auckland, Hauraki, Bay of Plenty, Wellington, North Canterbury, and Otago); final classifications subject to appeal (Porirua-Paraparaumu), and reclassifications subject to appeal (Tauranga Harbour). The Council's opinion was that these classifications were not in accord with the Cooke's decision.
[c] This was a reclassification of a part of the Kaituna-Rotorua classification. Appeals were outstanding when Cooke 's decision was published, but it was not withdrawn: being an upgrading of classes it was thought to be in accord with the decision.
[d] Classification done after Cooke 's decision. The Hawke Bay classification is the subject of an appeal to the High Court, by the NZ Underwater Association, on a point of law.
[e] Reclassification, done after Cooke's decision.
[f] Finalised before Cooke 's decision, because there were no appeals.

Source: Evidence presented by W.R. Howie to Cooke, J., and P Prendergast (*pers. comm.*)

APPENDIX 1: SAMPLE SUMMARY STATISTICS

Take the case where we have n data, represented by X_i: i = 1, ... ,n. Then the following statistics may be calculated, where Σ implies summation from 1 to n.

Arithmetic mean $\bar{X} = \Sigma(X_i)/n$

Median The middle value: half being larger and half being smaller. If n is even, the median is the average of the two middle values.

Percentile Written as r%ile, where r is a number from 0 to 100. The r%ile is that value of X for which r% of the data is of smaller magnitude. The 100%ile is the maximum; the 50%ile is the median.

Geometric mean $\bar{X}_g = \sqrt[n]{X_1 X_2 \ldots X_n}$

Mean log $\overline{(\ln X)} = [\Sigma(\ln X_i)]/n = \ln(\bar{X}_g)$

The mean log is not used in water quality criteria work. It is presented here to clarify the meaning of the ambiguous phrase "log mean" introduced in NTAC (1968): their "log mean" is actually the geometric mean, whereas, at face value, one might expect it to be the logarithm of the arithmetic mean. Note that the geometric mean is the antilog of the mean log. It is far, far better to stick to the term geometric mean, and ignore "mean log" or "log mean" altogether.

APPENDIX 2: INTERPRETATION OF SIGNIFICANCE TESTS

Considerable care must be exercised when comparing the results of significance tests. Let's assume that two investigators use the same methods to examine whether there is a relationship between indicator density and disease risk. They each obtain a set of data using these same methods and use the same statistical test to measure the strength of whatever relationship exists in each of the two populations they have sampled. The test asks whether the sample linear correlation coefficient (r) is greater than zero merely by chance. Because there is always a chance of reaching the wrong conclusion, a 95% confidence level is used, i.e., the significance level is $\alpha = 0.05$. This means that if the slope is concluded to be "significant" there would be at most a 5% chance of being wrong. The test is entirely equivalent to a test on the slope of the regression line. Because they cannot conceive that increasing indicator densities will result in a decrease in the disease risk (i.e., the slope cannot be negative), a one-sided test is used, as follows (e.g., Zar 1984):

$$\text{if } r\sqrt{\frac{n-2}{1-r^2}} > t_{\alpha,n-2} \text{ then conclude that a significant correlation exists.}$$

where $t_{\alpha,n-2}$ is read from a standard t-table for n samples. The first investigator finds that the slope is significantly greater than zero, the second finds that it isn't. What should be interpreted from this?

You can't interpret the results until you are given two extra pieces of information. You need to know both the p-value achieved by each investigator, *and* the number of samples they used.

The p-value is the smallest significance level that would have resulted in concluding "significance". So it is the actual probability of being wrong if it is concluded that there is a positive correlation. In our hypothetical case one investigator got $p > .05$ and the other got $p < .05$. But if one actually achieved a p-value of .045 and the other got .055 the results could hardly be said to be much different.

But this information alone is not enough. The left-hand-side of the inequality above will tend to increase with n (in individual cases it varies also with r of course, but on average r is independent of sample size). That is, the ability to detect a given slope increases as the sample size increases. So we can really only consider that the two investigators got different results *if they collected the same number of samples.* In statistical parlance, the power of the test increases with sample size (or, equivalently, the Type II error risk of falsely accepting the null hypothesis decreases with sample size).

It is an easy matter to show from the above formula that a correlation coefficient will be found to be significantly greater than zero if

$$r > \frac{t_{\alpha,n-2}}{\sqrt{n-2+t^2_{\alpha,n-2}}}$$

so that at the 5% significance level the following pairings of n and r will lead to inferring a significant correlation:

n	|r|
10	>0.549
50	>0.235
100	>0.165
500	>0.074
1000	>0.052

This shows clearly that the more samples one takes, the more likely it is that the correlation will be found to be significant. So, for example, the different conclusions reached by the two investigators could occur if if the both investigators got a correlation coefficient of 0.20, but the first had 100 samples and the second only 20. Tables such as this appear in some texts (e.g., Zar 1984).

APPENDIX 3: COMPLIANCE WITH PERCENTILE STANDARDS

In this Appendix I attempt to show that 5 samples in a assessment period gives poor performance in the examination of compliance. Indeed, an optimal performance is usually attained at around 50 samples (unpublished manuscript). That many samples is rather impractical in many cases, but at least by extending the length of the assessment period (it doesn't have to be 30 days) more can be obtained.

A percentile standard requires that a numerical limit should not be exceeded for more than a given percentage of a stated assessment period. The usual approach to testing compliance with such a standard is to collect random samples, to make an unbiased estimate of the percentile. If the statistical distribution of the population from which the sample has been drawn is known, then a parametric estimate of the population (true) percentile can be made. However, samples are typically so small that we lack any satisfactory statistical power to identify the population distribution at all well. In this case, the sample percentile is used as the unbiased estimate of the true percentile, and we must use nonparametric statistical procedures (performing tests on the ranks of the data).

The estimated percentile can then be compared to the limit to see if it was exceeded. This is of course the statistical procedure of "hypothesis testing".

But it may be objected that because the estimate is unbiased, there is a 50-50 chance that the true percentile was lower than the estimate. If the estimated percentile was just above the limit, so that it was concluded that the standard had been breached, then there is a 50% chance that the inference is wrong (in hypothesis testing this is the Type I error risk, denoted by α) . The usual way to minimise this risk is to compare the upper confidence limit of the sample percentile with the limit. But what is not usually pointed out is that in so doing one magnifies the risk of inferring that the standard hasn't been breached when in fact it has (this is the Type II error risk, denoted by β). Having fixed α, the only way to reduce β is to take more samples. So designing the necessary sampling programme, i.e., calculating the required sample size, is necessarily an exercise in balancing these two risks.

The statistical calculations for all this can get a bit messy. If you really do know that your data are distributed normally or lognormally (and that's highly debatable for microbes), it can be done, and it will be the most accurate approach. But in the usual case of not so knowing, one seems to have to resort to inaccurate nonparametric confidence intervals. And for upper percentiles, lots of data.

There is an easy, and simple, way out. It abandons the idea of estimating the percentile at all, and instead focusses on the *compliance percentage,* i.e., the true percentage of the compliance assessment period in which the limit is not breached. Compliance is to be judged by a rule which says: if we get no more than e exceedances in n samples we will conclude that the standard has been met, if we get more than e exceedances in n samples we will conclude that the standard has been breached. The beauty of the technique is that because samples were taken at random, for any true compliance percentage we can calculate the probability of getting up to e exceedances, and can hence quantify both error risks for any true compliance percentage. This can, for small sample sizes ($\leq$20) be done by hand and a binomial probability table, but it is simpler to do it on computer. The theory is given in McBride and Pridmore 1987 - for effluent compliance, but the same logic applies to receiving water compliance. The program is OCCAM (McBride 1987).

It is best demonstrated by example. Suppose we take 5 samples to test for compliance with a median (i.e., 50%ile) standard. Comparing the sample median to the limit is equivalent to adopting a (n=5,e=2) rule: we will infer compliance if we get no more than 2 exceedances in 5 samples. If we plot the probability of inferring compliance

(i.e., of getting no more than 2 exceedances) against the true compliance percentage, and draw the median line, we can immediately see both errors (the curve so obtained is known as an "operating characteristic curve"). These are shown on Fig. 1. If, for example, the true compliance percentage was 40%, there is a more than 30% risk of getting no more than two exceedances and so wrongly inferring compliance (i.e., $\beta>0.3$). Alternatively, if the true compliance percentage was 70%, there is a sizeable risk (about $\alpha=15\%$) of inferring breach. If the true compliance percentage was very close to 50%, then the error risks are both 50% also, as expected for this rule.

If these risks are unacceptable, and they would be to many, there are two things you can do. First you can take more samples, still using the sample median. This steepens the operating characteristic curve, and so narrows the window of true compliance where significant error risks occur. For example, see the (11,5) curve on Fig. 1. Secondly, you can push the operating characteristic curve further into the breach or compliance region by raising or lowering the number of allowable exceedances, e. This is shown on Fig. 2 where (11,6) and (11,4) operating characteristic curves are compared. By comparing these with the (5,2) curve you can see the advantage of collecting more samples.

Finally, you can calculate how many samples to get a given degree of protection for assessing compliance with a maximum standard. In such a case, there is no Type I error risk: ignoring analytical error, if you ever get an exceedance then you can conclude absolutely that the standard has been breached. So if you set the Type II error risk at, say, 5%, you can calculate the number of samples needed to detect a given true compliance percentage - as in Fig. 3. On this Figure we see the common pattern (also seen on such curves for assessing compliance with lesser percentiles) that around 50 samples the curve steepens upward so that little is gained by more sampling.

The figure shows that the true compliance that can be detected for monitoring a maximum standard. The risk of failing to detect a breach of the maximum is $\beta = 0.05$. So for 5 samples the true compliance would have to drop to 55% before one could be 95% certain of getting one exceedance. For 50 samples the true compliance would have to drop only to about 94%. The more samples, the better the protection.

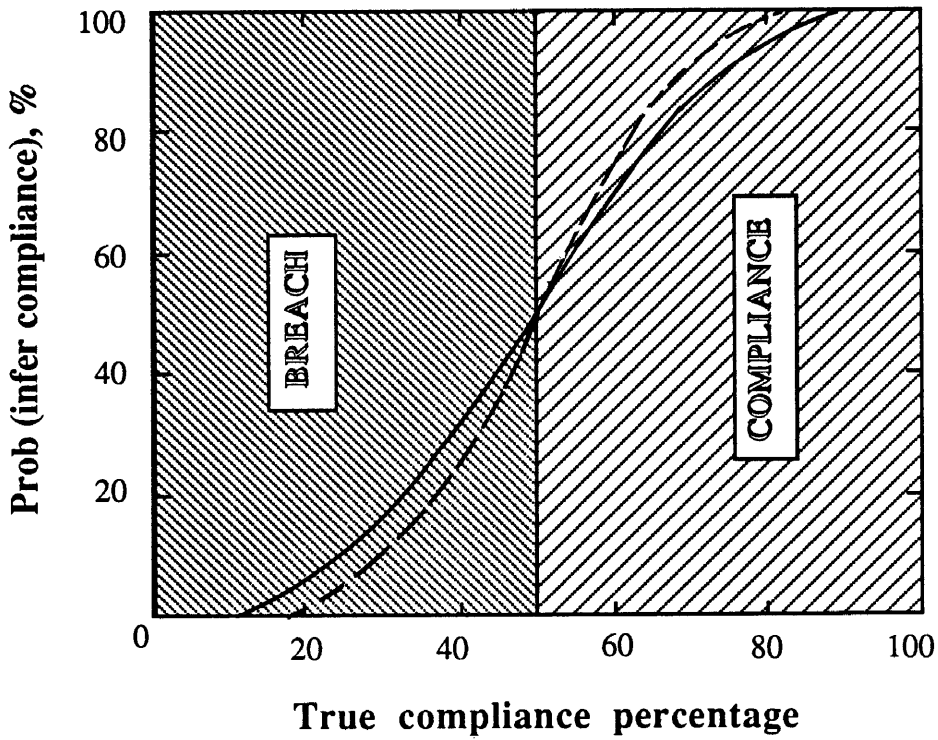

Figure 1. Operating characteristic curve for two unbiased compliance rules
(——— n = 5 samples, e = 2 allowable exceedances)
(— — n = 11 samples, e = 5 allowable exceedances)

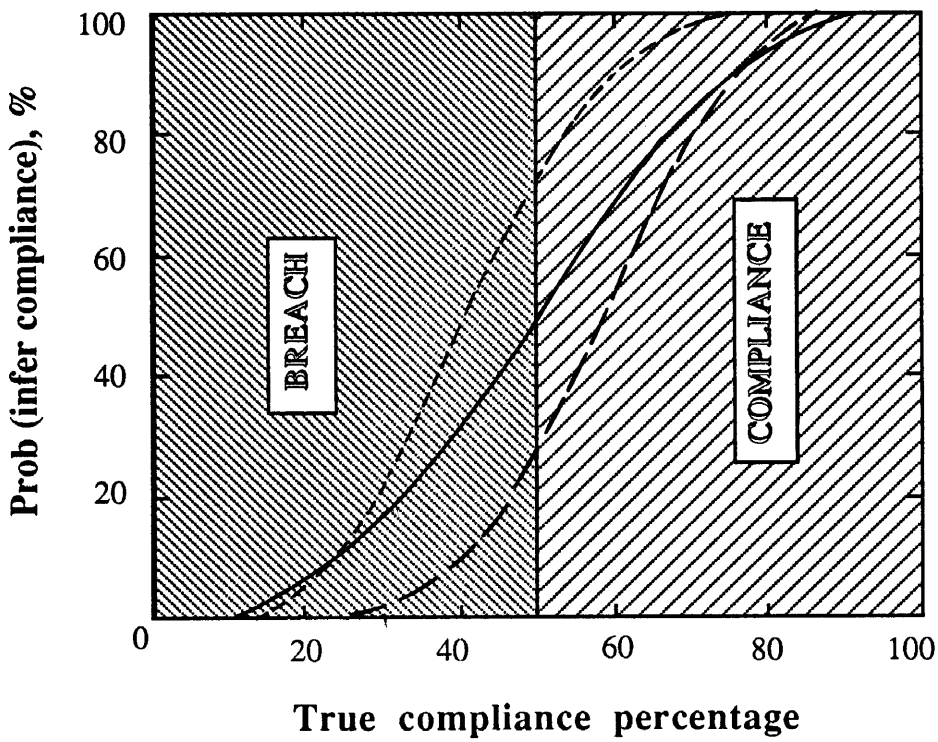

Figure 2. Operating characteristic curve for two compliance rules
(——— n = 5 samples, e = 2 allowable exceedances)
(— — n = 11 samples, e = 4 allowable exceedances)
(- - - - - - n = 11 samples, e = 6 allowable exceedances)

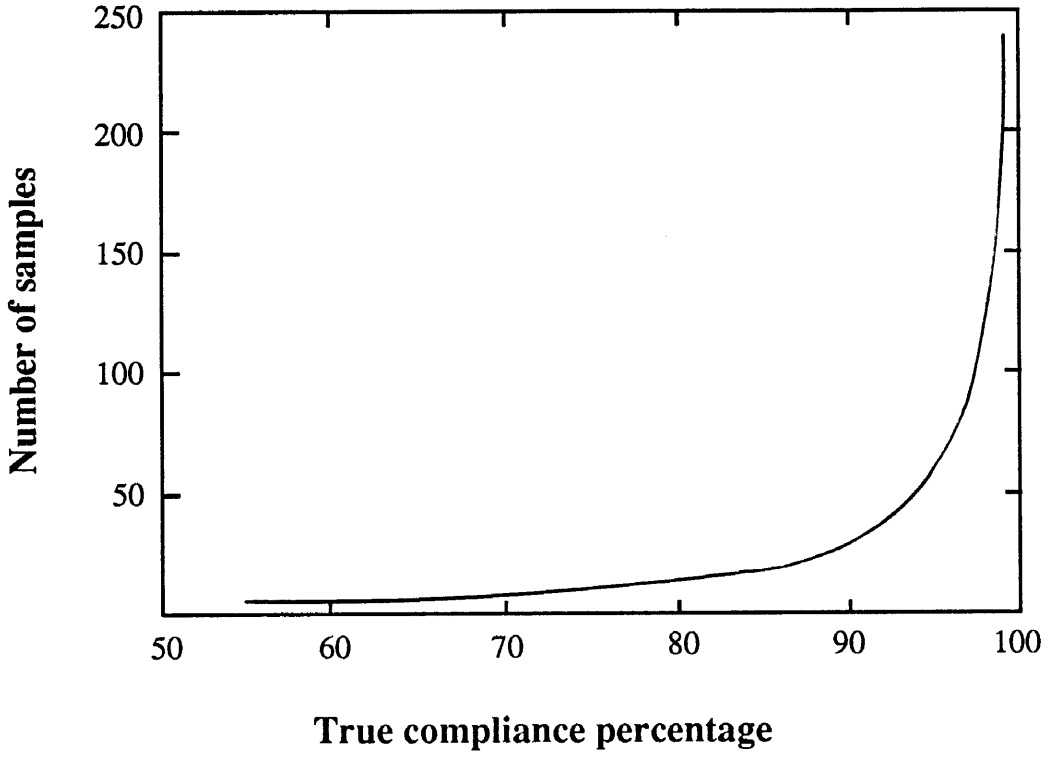

Figure 3. Number of samples needed to monitor compliance with a maximum: Type II error risk $\beta = 0.05$.

WATER QUALITY CENTRE

100 Aurora Tce,
Hamilton.
Postal Address:
PO Box 11-115
Hamilton, N.Z.
Fax: (071) 560-151
Telephone: (071) 567-026

DIVISION OF WATER SCIENCES
DSIR