Institute for Environment and Health

A REVIEW OF THE HEALTH EFFECTS OF SEA BATHING WATER

IEH Web Report W2

NOVEMBER 2000

The Institute for Environment and Health was established by the Medical Research Council at the University of Leicester in 1993. The Institute is partly funded by the Department of the Environment, Transport and the Regions, the Department of Health and other government departments and agencies by way of specific research and consultancy contracts.

This report is based on an unpublished client report to the Department of the Environment, Transport and the Regions issued in November 2000.

The views expressed here do not necessarily represent those of any government department or agency.

Written by MA Mugglestone, ED Stutt and L Rushton

Edited by JM Emeny

Acknowledgements: This work was peer-reviewed by members of the Public Health Laboratory Service Advisory Committee on Water and the Environment. We are grateful for the comments they provided.

Please cite as:
IEH (2000) *A Review of the Health Effects of Sea Bathing Water* Written by MA Mugglestone, ED Stutt and L Rushton (Web Report W2), Leicester, UK, Institute for Environment and Health (at http://www.le.ac.uk/ieh/webpub/webpub.html posted November 2000)

MRC Institute for Environment and Health
University of Leicester
94 Regent Road
Leicester
LE1 7DD
UK

Contents

Executive summary

BACKGROUND

The Bathing Water Directive (76/160/EEC) adopted by the Council of the European Communities (EC) *inter alia* sets mandatory and guideline microbiological standards for total coliforms and faecal coliforms, and guideline standards only for faecal streptococci. Data from a series of four UK epidemiological randomised controlled trials have been used to develop dose-response models which suggest that adverse health effects may occur at levels lower than the current standards. The World Health Organization (WHO) has produced draft guidelines for safe recreational water environments, Chapter 4 of which (microbiological aspects of water quality) places considerable weight on the UK randomised controlled trials. The Institute for Environment and Health (IEH) was asked by the Department of the Environment, Transport and the Regions (DETR) to carry out a critical review and evaluation of the literature on the health risks associated with recreational sea water contact.

The aims of the review were:

- **to evaluate critically all aspects of the design, conduct, analysis and interpretation of results from the UK randomised controlled trials, including data quality, the appropriateness of the statistical methodology and any potential biases;**
- **to discuss the UK randomised trials in the context of other studies reviewed by the WHO and to comment on their use of the evidence to formulate a risk assessment; and**
- **to make recommendations for additional work which might clarify some of the concerns identified.**

The review does not consider the clinical significance and public health impact of the results of the UK trials nor of the results of the WHO risk assessment.

These issues will need to be addressed by the DETR in order to obtain a complete understanding of the potential implications of the WHO draft guidelines.

THE WHO GUIDELINE DOCUMENT

Issues of general concern in the study of the health effects of sea bathing water are highlighted in this section and comments are made on the weight given to these by the WHO.

Indicator organisms are used as proxy measurements of the pathogens which actually cause ill health and varying relationships between indicators have been found in different studies. Concentrations of indicator organisms vary considerably both temporally and spatially, depending on environmental factors and local conditions. Standard techniques exist for the microbiological analysis of water samples. However, imprecise estimations of indicator counts can occur due to the variations highlighted above, the imprecise nature of analytical methods of bacterial enumeration, and interlaboratory variability in sample analysis.

Potential adverse health outcomes include gastrointestinal illness, eye infections, skin complaints, ear, nose and throat infections and respiratory disease. Most studies measure self-reported gastrointestinal symptoms.

Most of the articles reviewed by the WHO describe prospective or retrospective cohort studies. A different type of study is the randomised controlled trial, in which subjects are randomly assigned to bathing and non-bathing groups. Information about subjects recruited to studies of the effects of sea bathing is generally collected by interview or postal questionnaire (whether cohort or randomised controlled trial). Both types of study can thus potentially suffer from recall bias. Cohort studies may be limited in that, unlike randomised controlled trials, exposure status (bather/non-bather) is not controlled by the investigators, and subjects who choose to bathe or not

may differ demographically. Differences in composition may thus be smaller between bathing and non-bathing groups in a randomised controlled trial than in a cohort study. Statistical methodology exists to adjust for these differences through identification of so-called confounding factors, and such methods should be incorporated into the analysis of any study. Randomised controlled trials are limited in the range of exposures to which bathers may be ethically subjected, and may also be restricted to studying the effects on healthy adults only. Randomised controlled trials of sea bathing water are not as precise as randomised controlled drug trials. Dose-response models can be used to quantify relationships between health outcomes and degree of exposure in both cohort studies and randomised controlled trials.

Faecal streptococci appears to be an appropriate bacterial indicator of the risks associated with unknown viral pathogens. Studies assessing relationships between water quality and ill health should be based on water samples obtained at the location and time of exposure, and quality control analyses should be undertaken. Gastroenteritis is the most frequently reported adverse health effect resulting from sea bathing and the incubation period is typically no more than four days. The definition of gastroenteritis often differs between studies because of the reliance on self-reported symptoms.

The WHO review focuses on Kay *et al.*'s (1994) dose-response model because it is derived from randomised controlled trials, links gastroenteritis to counts of faecal streptococci made at the time and location of exposure, and because data on potential confounding variables were recorded.

THE UK RANDOMISED CONTROLLED TRIALS

Four randomised controlled trials and ten prospective cohort studies were conducted in the UK between 1989 and 1992. In the randomised controlled trials subjects were recruited from the local population and randomised to bathing or non-bathing groups. Bathers spent at least 10 minutes in the water and were asked to immerse their heads

at least three times. Exposure was measured using recognised indicators (total coliforms, faecal coliforms, faecal streptococci, total staphylococci and *Pseudomonas aeruginosa*). Exposure measurements were obtained at the time and location of exposure and at three different depths (surf, mid and chest). Data on adverse health effects were collected by four questionnaires (initial, exposure-day, one week post-exposure and three-weeks post-exposure). Subjects provided information about flu/cold, chest, ear/eye, gut, skin and other symptoms experienced. Subjects were examined by a physician at the time of the initial and one-week post-exposure questionnaires and biological samples (ear and throat swabs and faecal specimens) were also collected. Kay *et al.* (1994) relate gastrointestinal symptoms reported in the randomised controlled trials to measurements of exposure. Occurrence of gastrointestinal symptoms was chosen for detailed analysis because these symptoms had been reported most frequently in other studies. The use of self-reported gastrointestinal symptoms may have lead to misclassification of the health outcome. The definition used by Kay *et al.* (1994) was slightly less stringent than those of other studies and the observation period for gastrointestinal symptoms was 21 days. Almost two-thirds of gastroenteritis incidences occurred after seven days. It is surprising that the researchers considered these post-seven-day reports as being associated with the measured bathing exposure, as the incubation period is typically between two and seven days. The rates for bathers and non-bathers separately were not reported. Only faecal streptococci measured at chest depth showed a statistically significant relationship with gastroenteritis. Analyses of the data from medical examinations and bodily samples failed to show any relationship between clinically detectable adverse health outcomes and exposure but Kay *et al.* (1994) failed to mention these results. Most causes of gastrointestinal symptoms are, however, undetectable by common clinical methods.

The UK randomised controlled trials and the results reported by Kay *et al.* (1994) display some positive features but there are many serious weaknesses that render the results unreliable. The main strengths of Kay *et al.*'s (1994)

dose-response model are that it links incidence of gastroenteritis to faecal streptococci counts, and that it is based on randomised controlled trials.

Overall, sources of bias in the design, conduct and analysis of the UK randomised controlled trials are expected to have led to serious over-estimation of the risks of sea bathing. Bias would have affected estimates of the baseline risks of experiencing gastroenteritis (that is, risks to non-bathers) as well as risks to bathers and relative risks (bathers compared with non-bathers). The major sources of bias are:

- **restriction to healthy adult volunteers;**
- **self-reporting of symptoms;**
- **the length of the recall period in questionnaires;**
- **the definition of gastroenteritis;**
- **the three-week post-exposure observation period;**
- **repeated statistical testing;**
- **model selection;**
- **recruitment methods; and**
- **timing of exposure-day interviews.**

The only source of bias recognised by Kay *et al.* (1994) was the one source that would be expected to lead to under-estimation of risks (restriction to healthy adult volunteers).

Serious weaknesses in Kay *et al.*'s (1994) statistical analysis throw the validity and reliability of their dose-response model into doubt. The weaknesses include:

- **over-stringent rejection of potential confounders on which data had been collected;**

- **further reduction of the set of identified confounders using a poorly described composite variable;**
- **failure to examine the effects of other potential confounders such as additional bathing activity, holidays/business trips in the UK and abroad, and awareness of beach maintenance/pollution at any stage in the analysis;**
- **failure to control for any confounders at all in the final dose-response model;**
- **reduction in the effective sample size due to exclusion of non-bathers;**
- **insufficient consideration of the most appropriate transformation of faecal streptococci counts to use as a continuous explanatory variable in dose-response modelling;**
- **invalid methodology for estimating the threshold of risk;**
- **failure to present estimates of uncertainty (standard errors or confidence intervals) for the fitted model;**
- **expression of the fitted model in terms of excess risk; and**
- **failure to analyse for different latency periods.**

Given the extensive criticisms of Kay *et al.* (1994) detailed above it is surprising that the article was regarded as suitable for publication in the form in which it appeared. All of the weaknesses in the statistical analysis could be addressed by appropriate re-analysis, and several sources of bias could be eliminated (or at least reduced) at the same time.

THE UK RANDOMISED CONTROLLED TRIALS IN CONTEXT WITH OTHER STUDIES

Over a dozen major epidemiological studies have been conducted in marine waters, the majority being prospective cohort studies. Many studies, like the UK randomised controlled trials, show a general increase in adverse health symptoms with decreasing

water quality, although this is not a consistent finding across all studies. Considerable variations in risk estimates and incidence rates associated with both non-bathers and bathers are reported in these studies, with very few indicating a consistent and strongly increasing dose-response relationship. The WHO guideline document reports unadjusted relative risks and associated confidence intervals, that is, it ignores any potential confounders.

The UK randomised controlled trials appear to be less subject to non-differential bias than other major studies in terms of the precision with which exposure was controlled and measured but they are subject to many other sources of bias. Kay *et al.* (1994) report strikingly higher incidence rates of gastroenteritis than do other studies in both bathers and non-bathers and the overall position of their dose-response curve is higher up the vertical (risk) axis than curves for other studies. This could be explained by bias in the estimation of baseline risks (to non-bathers, or to bathers in the lowest category of pollution). Kay *et al.*'s (1994) dose-response curve is also steeper than the curves from the other studies. This could be explained by inadequate control for confounders.

The WHO guideline document compares estimated thresholds of risk from different studies. The threshold implied by Kay *et al.*'s (1994) dose-response model is a consequence of the way the model was fitted (the model forces zero risk below the threshold). The meaning of the term threshold when applied to the other studies refers to the level of water quality at which a statistically significant elevation in risk (relative to non-bathers, or to bathers in the least polluted water, depending on the study) is observed. Neither use of the term threshold necessarily implies a clinically significant risk to health, nor a level of faecal contamination that necessarily requires intervention. Kay *et al.*'s (1994) dose-response curve is presented without estimates of the uncertainty of the fitted model (standard errors or confidence intervals) which hinders comparison with curves from other studies. The small sample size of the UK

randomised controlled trials could result in larger standard errors and confidence intervals than in many of the other studies.

Several dose-response models reviewed in the WHO guideline document were derived from the UK randomised controlled trials and are subject to the same or similar limitations as Kay *et al.*'s (1994) model. The review conducted by the WHO is selective in that results from studies that have not shown associations between adverse health effects and bathing water quality have not been reproduced. The WHO review is also misleading in that it implies that self-reported symptoms in the UK randomised controlled trials were confirmed by medical examination. The relative risks and confidence intervals presented by the WHO are of limited value because they are unadjusted for confounders.

Only one non-UK study (Cabelli *et al.*, 1982) presents a dose-response relationship between gastroenteritis and bathing water quality. Of all the studies reviewed by the WHO and those that have appeared subsequently, the UK randomised controlled trials probably provide the most precise estimates of water quality and exposure, and they have the potential to provide highly acceptable estimates of adverse health effects. Without re-analysis of the randomised controlled trials, however, the higher risk estimates reported by Kay *et al.* (1994), including baseline risks to non-bathers and relative risks for bathers in waters of different quality, cannot be regarded as reliable. These features are, instead, likely to be the result of seriously biased estimates of gastrointestinal symptoms and inadequate statistical analysis.

USE OF THE RESULTS OF THE UK RANDOMISED CONTROLLED TRIALS IN RISK ASSESSMENT

Several issues require consideration in the development of an ideal risk assessment relating health effects to environmental exposure, namely: characterisation of exposure; characterisation of health effects; existence of a dose-response relationship

quantifying the relationship between the level of exposure and health effects; characterisation of the degree of uncertainty in risk estimates by means of standard errors or confidence intervals; characterisation of all sources of variation in risk and presentation of the results of risk assessment at a disaggregated level; and analysis of the sensitivity of results to model assumptions.

The basic elements of the risk assessment strategy used in the WHO guideline document are:

- characterisation of exposure using faecal streptococci as an indicator of water quality;
- characterisation of health effects in terms of gastrointestinal symptoms;
- adoption of Kay *et al.*'s (1994) dose-response model quantifying the relationship between gastroenteritis and water quality;
- an assumption that $\log_{10}$ faecal streptococci counts are normally distributed with a standard deviation of 0.8103;
- definition of the criteria for compliance with microbiological standards to be that 95% of water samples will result in faecal streptococci counts equal to or less than the standard;
- calculation of the mean $\log_{10}$ faecal streptococci count (and hence the 95 percentile to be proposed as a standard or guideline value) required for the expected proportion of exposures that lead to gastroenteritis to be less than a chosen target value; and
- comparison of guideline values for faecal streptococci counts corresponding to different choices for the target value.

The reliance on faecal streptococci as an indicator of faecal pollution and gastroenteritis as an adverse health effect can be regarded as appropriate but there are serious concerns regarding the rest of the approach to risk assessment taken in the WHO guideline document. The concerns include:

- reliance on Kay *et al.*'s (1994) dose-response model, which is subject to bias, inadequate control for confounders and other limitations as discussed earlier;
- failure to characterise the degree of uncertainty in risk estimates;
- failure to characterise all sources of variation in risks and to present the results of risk assessment in disaggregated form; and
- failure to undertake any form of sensitivity analysis.

The expected impact of the reliance on Kay *et al.*'s (1994) dose-response model is to over-estimate the risk of gastroenteritis at a given exposure level. Thus, the proposed guideline values may be lower than necessary.

Major improvements in the WHO risk assessment (and hence the proposed guideline values) could be obtained by incorporating a new dose-response model derived from appropriate re-analysis of the UK randomised controlled trial data. Not all sources of bias in the UK randomised controlled trials can be eliminated by re-analysis. Interpretation of recalculated guideline values must, therefore, incorporate consideration of the expected effects of bias.

The other limitations of the WHO risk assessment must also be addressed in order to arrive at reliable guideline values. Characterisation of uncertainty and variability in risk estimates, as well as sensitivity analysis could be achieved by, for example, Monte Carlo simulation, although other equally valid statistical techniques (for example, Bayesian inference) could be used. Uncertainty relates to the estimated dose-response relationship and the estimated standard deviation of $\log_{10}$ faecal streptococci counts, whilst variability relates to the effects of confounders (age, sex, etc.) and other variables (for example, geographical location). Assumptions of independence between risks for individuals in the same family group and between different bathing episodes for a single individual, as well as the effects of relying on Kay

_et al._'s (1994) dose-response model rather than, say, Cabelli _et al._'s (1982) dose-response model should be examined as part of sensitivity analysis.

RECOMMENDATIONS

This review has identified two major areas of concern in the analysis presented in the WHO guideline document, namely: the statistical analysis of the UK randomised controlled trials to yield Kay _et al._'s (1994) dose-response model linking gastrointestinal symptoms to faecal streptococci counts; and the WHO risk assessment strategy used to derive guideline microbiological standards. Both areas of concern throw into doubt the validity and reliability of the proposed guideline values, and both areas require re-analysis in order to obtain satisfactory estimates of guideline values.

The minimum requirements for re-analysis of the UK randomised controlled trial data are as follows. The incidence of gastrointestinal symptoms (in both bathers and non-bathers) in relation to faecal streptococci counts and other factors should be examined using logistic regression models. The response variable (incidence of gastrointestinal symptoms) should be defined to reflect both the severity of symptoms used to define gastroenteritis in other studies (for example, Cabelli _et al._, 1982) and the expected latency period. Sensible choices for the latency period would be:

- **symptoms recorded within seven days of exposure;**
- **symptoms recorded between eight and 14 days after exposure; and**
- **symptoms recorded between 15 and 21 days after exposure.**

Models for different response variables should be compared with each other and with results from other studies.

The following explanatory variables should be included in the models.

- **Exposure status (bather/non-bather) represented as an indicator variable. Inclusion of such a term would allow the effects of sea water alone (irrespective of the degree of pollution) to be investigated. This aspect could be tackled by fitting the indicator variable to estimate the average effect of contact with sea water, plus an interaction between the indicator and faecal streptococci counts (see below) to estimate the difference between the average effect and the effect at different levels of water quality.**
- **Geographical location (trial site) represented as a factor or a random effect.**
- **Potential confounders:**

  - age;

  - sex;

  - frequency of diarrhoea;

  - illness in the four weeks preceding exposure day that lasted for more than 24 hours;

  - consumption of foods associated with gastroenteritis;

  - illnesses in the household between exposure and the end of symptom recording;

  - additional bathing in the period from three days before exposure to the end of symptom recording;

  - awareness of beach maintenance in the UK;

  - awareness of pollution;

  - awareness of news/media coverage of the trials;

  - holidays/business trips in the UK and abroad in the period from four weeks before exposure to the end of symptom recording;

  - duration of exposure; and

  - ingestion of water.

Representation of one or more confounders as a single composite variable should be avoided.

- **Faecal streptococci counts represented as a continuous variable.**

Terms that are not measurements of exposure should be fitted first, so that variability in risks due to these sources can be removed before examining the effects of water quality, etc. The analysis of deviance table should be examined to compare the strengths of the effects of different confounding variables and exposure. Standard errors or confidence intervals for risk estimates, odds ratios, or estimated effects of individual explanatory variables should be reported. Goodness-of-fit of the models should also be discussed.

The analysis should include an assessment of the form of the continuous variable representing faecal streptococci counts that gives the best fit. It would also be advisable to explore the possibility of interactions between faecal streptococci counts and factors such as geographical location. The evidence for a non-linear relationship between log-odds of gastrointestinal symptoms and faecal streptococci counts could be explored by fitting a quadratic term in faecal streptococci. Fitted models obtained using the methods outlined above would yield estimates of risks of experiencing gastroenteritis for both bathers and non-bathers, at a given level of water quality and at given levels of other factors in the models. These estimates, and the corresponding standard errors or confidence intervals, could then be fed into the WHO risk assessment.

Re-analysis according to these specifications will ameliorate bias due to the definition of gastroenteritis, the three-week observation period, repeated statistical testing and model-selection bias. Additional sources of bias that are inherent in the study design but cannot be quantified by re-analysis are:

- **restriction to healthy adult volunteers;**

- **self-reporting of symptoms;**
- **length of recall period;**
- **recruitment methods; and**
- **timing of exposure-day interviews.**

The interpretation of the results obtained by re-analysis should attempt to incorporate qualitative statements regarding the expected effects of these sources of bias.

The WHO risk assessment strategy requires modification to incorporate uncertainty, variability, and sensitivity analysis. These issues could be addressed using Monte Carlo simulation or any other appropriate statistical technique.

The analysis of uncertainty in risk estimates should address both the uncertainty attached to the re-fitted dose-response model, and uncertainty due to the unknown distribution of $\log_{10}$ faecal streptococci counts.

The major sources of variation in risk that should be examined and quantified are those due to potential confounders and geographical locations. After re-analysis of the UK randomised controlled trial data, it should be possible to present guideline values for more representative groups of individuals (relative to the entire population) than are considered by the current proposed guidelines. In this way population-attributable risk could be estimated. Factors that could be taken into account are age, sex, etc. The results should also be presented in disaggregated form to allow the effects of different sources of variation in risk to be compared.

Sensitivity analysis should examine the effects of assumptions made during risk assessment. These include independence of risks between different exposure occasions, and between different individuals in a family group. The sensitivity of the guideline values to the assumed standard deviation for $\log_{10}$ faecal streptococci counts could also be investigated. The value used is an average over several EC Member States. Values for individual Member States could be used instead. Finally, the sensitivity of the guideline values to the UK randomised controlled trial data could be explored by extending the risk assessment to use dose-response models from other studies, for example, Cabelli *et al.* (1982).

1 Background and introduction

The Bathing Water Directive (76/160/EEC) adopted by the Council of the European Communities (EC) *inter alia* sets mandatory and guideline microbiological standards for total coliforms and faecal coliforms, and guideline standards only for faecal streptococci. Concern was expressed in the late 1980s as to whether these standards provided sufficient protection to health. A series of UK epidemiological studies commissioned by the UK government came to the broad conclusion that the risk of contracting serious illness from sea bathing was negligible and that existing mandatory standards provided adequate protection. Dose-response models have been developed using the data from randomised controlled trials conducted as part of these studies. The models suggest that adverse health effects may occur at levels lower than the current standards. The World Health Organization (WHO) has now produced draft guidelines for safe recreational water environments which review the evidence and which place considerable weight on the results from the UK randomised controlled trials. These guidelines are being considered by the EC Commission and Member States as part of the proposed future revision of the Bathing Water Directive. Reservations have been expressed about the design, data and statistical models on which the proposed guidelines are based. The Institute for Environment and Health (IEH) was asked by the Department of the Environment, Transport and the Regions (DETR) to carry out a critical review and evaluation of the literature on the health risks associated with recreational sea water contact. The structure of this review is as follows.

- Issues of general concern in the study of health effects of sea bathing water and comments on the weight given to these in Chapter 4 of the WHO guideline document are discussed in Section 2.
- A critical evaluation of all aspects of the design, conduct, analysis and interpretation of results from the UK randomised controlled trials, including data

quality, the appropriateness of the statistical methodology and potential biases is discussed in Section 3.

- The UK randomised controlled trials are discussed in the context of other studies considered in the WHO guideline document, together with some more recent studies and those excluded by the WHO in Section 4.
- The use of the evidence to formulate a risk assessment in the WHO guideline document is addressed in Section 5.
- Recommendations for additional work which might clarify some of the concerns identified are given in Section 6.

Each major section of the review is summarised at the end of the section.

The IEH review focuses on Kay *et al*. (1994) because the WHO guideline document relies on it so heavily. The review does not address the clinical significance and public health impact of the results of the UK trials nor of the results of the WHO risk assessment. These issues were not included in the IEH remit. Nevertheless, they will require consideration in order to obtain a complete understanding of the potential implications of the WHO draft guidelines.

A review of the literature on adverse health effects associated with recreational sea bathing was carried out by Prüss (1998). This review is used as the basis for Chapter 4 of the WHO draft guidelines. We have thus chosen to refer only to the WHO guideline document (WHO, 1998) throughout this report.

2 The WHO guideline document

2.1 OVERVIEW

This review relates specifically to WHO (1998, Chapter 4), which considers the evidence for relationships between the quality of recreational bathing water and adverse health effects and makes recommendations for microbiological standards to protect human health. Although WHO (1998) relates to both marine and fresh water environments, this review is restricted to the health effects of sea bathing. The structure of WHO (1998, Chapter 4) is as follows: epidemiological studies of the health effects of bathing water quality from across the world are reviewed; various different adverse health outcomes and bacterial indicator organisms used to measure exposure to faecal pollution are discussed; one particular article based on the UK randomised controlled trials (Kay *et al.,* 1994) is judged to provide the most accurate relationship between adverse health effects and water quality and is used to conduct a risk assessment from which guidelines for water quality standards are derived. The various aspects of WHO (1998, Chapter 4) are discussed in detail in the rest of Section 2.

2.2 EXPOSURES

Indicator organisms and pathogens

The potential of bathing water to cause adverse health effects is assessed by the measurement of bacterial indicator organisms which are associated with the presence of faecal contamination. Many pathogens (illness-causing organisms) present in faecal material cannot be measured directly and so indicator organisms are enumerated as proxies, with target species being those that are exclusively found in human and animal faeces. There are acknowledged difficulties and complexities in assessing relationships between pathogens and indicator organisms. However, WHO (1998) states that studies world-wide have demonstrated that ‘the rate of symptom acquisition for some diseases, correlates with the concentration of microbiological indicator species which

derive from the same sources as many of the pathogens'. All dose-response relationships are based on this assumption (see Section 2.4). Pathogens and indicator organisms can be present as a result of sewage contamination, the population using the bathing water, run-off from the surrounding land or local wildlife.

A number of different indicator organisms can be measured to assess bathing water quality. Those currently used by the EC for its quality standards are:

- total coliforms;
- faecal coliforms;
- faecal streptococci;
- salmonella; and
- enteroviruses.

Other bacteria used as indicator organisms for water quality are *Escherichia coli* (the dominant faecal coliform), *Pseudomonas aeruginosa* (a sewage indicator) and staphylococci. The coliforms are considered indicative of recent faecal pollution. Faecal streptococci have a longer survival time in marine waters (WHO, 1999)[1] and are considered to give a better representation of the survival of viral pathogens.

Comparison of different indicator organisms

Relationships between different indicators have rarely been examined but many studies have found relationships with health outcomes to be significant for one indicator but not for others (for example, Cabelli *et al.*, 1982; Kay *et al.*, 1994). The ratio between different indicator organisms is not constant but varies according to different circumstances. The ratio can be used to give an indication of the source of pollution (Olivieri, 1982). In addition to having a longer half-life, faecal streptococci are present in higher numbers in animal faeces than are faecal coliforms (Geldreich,

[1] WHO (1999) *Health-Based Monitoring of Recreational Waters: The Feasibility of a New Approach (The 'Annapolis Protocol')*. Protection of the Human Environment: Water, Sanitation and Health Series, Geneva, Switzerland, World Health Organization. Available [at November 2000] via http://www.who.int/water_sanitation_health/Water_quality/recreat.htm

1977). It is possible that different relationships exist between different pathogens and different indicators, and so it is useful to measure a variety of indicators for assessing the risks of adverse health outcomes. A few studies have demonstrated weak correlations between the numbers of different indicator bacteria. In particular, Nuzzi & Burhans (1997) found a correlation between enterococci and both total and faecal coliforms. Moreover, different classifications or labels often refer to very similar bacteria, facilitating their comparison. WHO (1998) groups together all studies regardless of the indicators used. It would be clearer to separate the studies into groups based on similar classifications.

The vast majority of faecal coliforms are *E. coli* (Figueras *et al.,* 1997). It would therefore be appropriate to group together faecal streptococci and *E. coli* for a comparison of measurements. In addition, the terms faecal streptococci and enterococci have been used interchangeably (Godfree *et al.,* 1997). Enterococci are bacteria belonging to the family Streptococcaeae which were formerly classified as part of the genus *Streptococcus* (Spraycar, 1995).

Indicator variability

Concentrations of indicator organisms vary considerably both temporally and spatially and there can be orders of magnitude difference within a few metres on the same beach and within a few hours at the same sampling point. In addition to distance and time from specific discharges, variations can be the result of:

- the degree and type of treatment applied to sewage;
- rainfall (following rainfall, stream waters can experience a 2- to 4-fold increase in $\log_{10}$ concentration even though discharge volume increases by only 1-2 orders of magnitude (Godfree *et al.,* 1997), leading to significantly elevated bacterial delivery to nearshore waters);
- wind;
- tides and currents;

- coastal physiography;
- amount of sunshine (sunlight is the most effective natural biocide); and
- high turbidity.

The effects of environmental factors and local conditions are discussed in detail in Serrano *et al.* (1998) and Jones & Obiri-Danso (1999). The general conclusion from this work is that these factors are correlated with indicator counts but that the correlations are too weak to be used in a predictive mode (Serrano *et al.,* 1998).

Most of the factors described above are adequately addressed by WHO (1998) and the recording of information relating to environmental factors and local conditions is recommended in WHO (1998, Chapter 11). Another potential source of variability is that due to analysis being carried out at different microbiological laboratories. Comparison of results from different laboratories for split samples seems to indicate appreciable inter-laboratory variability (Pike, 1994).

Microbiological analysis and quality control

The vast majority of epidemiological studies relating to the quality of sea bathing water use the standard technique of membrane filtration to estimate counts of indicator organisms. Although membrane filtration is thought to be the most accurate technique available, it is still thought to yield imprecise estimates of indicator organism density (Fleisher, 1990a). Coupled with large temporal and spatial variability in indicator numbers, this can lead to significant errors in the estimation of water quality. Fleisher (1990a,b) discusses this issue in detail, concluding that it is essential to make replicate determinations when conducting water quality surveys. Most studies only take one sample when making measurements of indicator density, except for limited quality control analysis.

Bather exposure

For the purpose of epidemiological studies (and any bather-relevant sampling regime) samples for indicator counts should be taken at the time and location when bathing actually occurs and at an appropriate depth. This is due to the large potential for indicator count variability discussed above. Results from studies using seasonal averages are likely to have extremely limited potential for risk assessment and should perhaps be reviewed separately from studies with more temporally and spatially relevant measurements. WHO (1998) states that 11 out of the 22 articles it reviewed attempted to relate adverse health outcomes to seasonal averages. None of these studies is used in the WHO risk assessment (see Section 5).

Choice of indicators

WHO (1998) concludes that faecal streptococci and enterococci are the indicators most frequently correlated with adverse health outcomes resulting from exposure to pathogens in sea bathing water. Of the articles reviewed in WHO (1998), the strongest relationship between exposure and gastroenteric symptoms was reported by Kay *et al.* (1994). It is generally accepted that faecal streptococci concentrations decline more slowly than those of faecal coliforms in nearshore marine waters (Godfree *et al.,* 1997). Godfree *et al.* (1997) review a number of epidemiological studies and conclude that faecal streptococci counts provide the best bacterial indicator of the risks associated with a probable viral aetiology. Figueras *et al.* (1997) state that faecal streptococci were included in the 1994 proposal for revised mandatory standards for the EC Bathing Water Directive in agreement with much research data.

2.3 HEALTH EFFECTS

Different health outcomes

Many possible adverse health outcomes could result from bathing in faecally contaminated marine water. WHO (1998) and the vast majority of epidemiological

studies address mild to moderate 'self-limiting' ailments and report that such health outcomes result from bathers ingesting faecally contaminated water. Gastrointestinal symptoms are most frequently reported and plausibly the most likely outcomes of exposure to sewage contamination given their faecal-oral route of transmission. Other potential outcomes include eye infections, skin complaints, ear, nose and throat infections and respiratory diseases (for example, acute febrile respiratory illness). WHO (1998) states that more serious illnesses such as hepatitis, cholera, typhoid fever, and meningo-encephalitis have been reported after exposure to heavily polluted recreational water but there are insufficient data to derive dose-response relationships or to estimate health risks for these outcomes.

Health outcome assessment

Assessment of gastroenteritis is difficult and there is a reliance on self-reported symptoms as opposed to clinical measurements because many of the pathogens are unknown or cannot be isolated from clinical samples. Direct comparison of studies may also be complicated by the use of different criteria to define health outcomes. Kay *et al.* (1994) define 'objective' gastroenteritis as 'any case of vomiting or diarrhoea plus any case of either indigestion or nausea accompanied by fever'. A number of US studies, including Haile *et al.* (1999) and Cabelli *et al.* (1982), use the term 'highly credible' gastrointestinal illness, which is deemed to occur if subjects report vomiting, or diarrhoea accompanied by fever, or stomach pain accompanied by fever.

Kay *et al.* (1994) base their results on incidences of gastroenteritis occurring up to 21 days after exposure to sea bathing water. Other studies indicate a much shorter incubation period for similar health outcomes. A freshwater study by Ferley *et al.* (1989) assumes latency to be three days for 'acute gastrointestinal disease' and Calderon *et al.* (1991) assume illnesses contracted within three days of bathing to be possibly water-related. Gastrointestinal syndromes in the general population can be caused by viruses, bacteria and parasites (Wheeler *et al.*, 1999). Viral gastroenteritis is

most common (Richman *et al*., 1997). The most common causes of viral gastroenteritis are the Norwalk agent and related viruses, which have incubation periods of 15-48 hours (Richman *et al.*, 1997). Other viral sources of gastroenteritis are rotaviruses, which have an incubation period of 2-4 days, and astroviruses, for which the incubation period is up to four days (Richman *et al.*, 1997). Adenoviruses can cause gastroenteritis (and febrile respiratory illness) and they have an incubation period of up to two weeks (Richman *et al.*, 1997). Enterotoxigenic *E. coli*, a bacterial cause of gastroenteritis, has a median incubation period of 42 hours (Dalton *et al.*, 1999), whilst parasitic diarrhoeas such as cryptosporidiosis have an incubation period of 7-10 days (Spraycar, 1995). The majority of identifiable aetiological agents that cause gastroenteritis therefore have incubation periods of less than seven days.

2.4 STUDY DESIGNS AND ANALYSES

Types of study

There are several ways in which epidemiological studies can be conducted. Most of the articles reviewed by WHO (1998) are based on prospective cohort studies. In this type of study, a healthy group of subjects (or cohort) is contacted at the beach and invited to participate. The subjects are asked to give information about recent bathing activity and about other factors, such as age. Some days after the beach interviews, the subjects are contacted again (typically by telephone) and asked to give information about their state of health, and about other factors such as bathing activity since the day of the beach interview. This type of study is prospective in that the follow-up period (between exposure and experience of adverse health effects) occurs after the time of the beach interviews.

Cohort studies can also be applied in a historical (or retrospective) mode. Two of the articles reviewed by WHO (1998) are based on studies of this type. In a historical cohort study, subjects are contacted at the beach and asked to give information about recent bathing activity and experience of adverse health effects. In this type of study the follow-up period occurs before the time of the beach interviews.

A fundamentally different type of study is a randomised controlled trial (or experiment) in which subjects are randomly assigned to bathing and non-bathing groups. Of the articles reviewed by WHO (1998), only three report the results of randomised controlled trials and all three in fact relate to a single series of trials conducted in the UK (Pike, 1994; Kay *et al.*, 1994; Fleisher *et al.*, 1996). Subjects were recruited in advance of these trials by means of various publicity drives. Additional recruitment took place at the beaches on the trial days. Subjects were followed up by personal interviews and postal questionnaires at specified intervals after exposure to obtain information about adverse health effects and other factors, such as additional bathing activity.

Advantages and disadvantages of different types of study

Each of the study designs outlined above has advantages and disadvantages. The most relevant factors in relation to the interpretation of study results and comparison between studies are described below.

Retrospective studies can suffer from recall bias, in that subjects who have experienced adverse health effects and those who have not may remember details about exposure and other factors at different rates. It may also be difficult to obtain accurate and detailed information about water quality at the time of exposure in retrospective studies.

Cohort studies are inherently limited in that exposure is not controlled by the investigators. Subjects who choose to bathe and those who do not may differ demographically (with respect to age, sex, etc.) or in other ways. The differences in composition between bathing and non-bathing groups may influence the rates at which the two groups experience adverse health effects. Providing that data documenting differences in group composition have been recorded and that differences between groups are not extreme, it is usually possible to adjust for their effects using

appropriate statistical methods. Variables that are a risk factor for the disease of interest, even in the absence of exposure, and are also associated with the exposure of interest are called confounding factors.

Randomised controlled trials have the following advantages over cohort studies.

- The random allocation of subjects to bathing and non-bathing groups should ensure that the composition of the two groups is roughly equivalent. Although differences between groups in a randomised controlled trial should be smaller than those in an observational study, it is still appropriate to collect data on possible confounding variables and to adjust for these during statistical analysis.
- The degree, duration and timing of exposure can be controlled. By timing measurements of water quality to coincide with exposure, the relationships between water quality and adverse health effects can be estimated more precisely than in an observational study.
- Particular groups of people involved in the trial (subjects, interviewers, etc.) can be made blind to certain aspects of the trial and this can help to eliminate certain sources of bias. For example, information about whether or not a particular subject was a bather or non-bather may be withheld from individuals conducting follow-up interviews.

Randomised controlled trials also have disadvantages in that, for ethical reasons, they are restricted to studying the effects on healthy adults of exposure at beaches that meet current standards. This may limit their use for prediction of risks of adverse health effects in the general population. In addition, the range of water qualities to which bathers can be exposed may not be as wide as that found in cohort studies or more generally. The above restrictions may affect the power of a trial to detect a statistically significant difference between rates of adverse health effects in bathers and non-bathers, and may limit the opportunities for investigating dose-response relationships (see below). On balance, this type of study might be expected to yield more precise estimates of the risks of experiencing adverse health outcomes in relation

to water quality than observational studies. However, it should be noted that randomised controlled trials of sea bathing water quality are not as precise as randomised controlled drug trials in which subjects are blind to the treatment they receive (subjects in epidemiological trials will know whether or not they are in the bathing group), and in which far greater control over subjects' behaviour can be exercised.

In addition to the problems described above, the method of recruiting subjects (at the beach or via publicity drives) might introduce bias into a study. This applies equally to cohort studies and randomised controlled trials.

A distinction is often made between non-differential and differential sources of bias in epidemiology. A non-differential bias arises when misclassification of health outcome or water quality is random with respect to exposure status (bather/non-bather). This type of bias usually leads to under-estimation of the risks of adverse health effects. A differential bias arises when misclassification of health outcome or water quality is linked to exposure status. For example, publicity about illnesses thought to be linked to faecal pollution might lead to over-reporting of symptoms in bathers and under-reporting in non-bathers. The overall effects of differential bias will depend on the direction and magnitude of the biases in the bathing and non-bathing groups (see Section 3.6).

Dose-response studies

Studies which examine relationships between different levels of exposure and adverse health effects can be used to estimate dose-response relationships. Such relationships are useful in that they quantify the risks of adverse health effects associated with given levels of water quality. Existence of a dose-response relationship adds weight to the evidence of a causal link between exposure and adverse health effects.

2.5 WEIGHT GIVEN TO THE BALANCE OF EVIDENCE

WHO (1998) gives greatest weight to the results of the UK randomised controlled trials reported by Kay *et al.* (1994). The rationale behind this choice is that this work combines: randomised controlled trials; a dose-response relationship linking gastrointestinal symptoms to concentrations of faecal streptococci measured at the time and place of exposure; and recording of data on potential confounding variables. The existence of a dose-response relationship allows WHO (1998) to derive guideline values for water quality using a probabilistic risk assessment model (see Section 5).

2.6 SUMMARY

Many different bacterial organisms can be used to measure faecal contamination of sea bathing water. The best indicator of risks of ill health due to unknown viral pathogens appears to be faecal streptococci. Indicator counts vary in space and time according to environmental factors and local conditions. Studies assessing the relationships between water quality and ill health should, therefore, be based on measurements of water quality obtained at the location and time of exposure. Membrane filtration is the best microbiological technique for enumerating bacteria but even this is inaccurate. Quality control analysis is, therefore, important.

Many different adverse health outcomes have been reported in relation to faecally contaminated sea bathing water. Gastroenteritis is the most frequently reported outcome. The definition of gastroenteritis in relation to self-reported symptoms and the latency period for infection require consideration.

Most of the studies reviewed in WHO (1998) are cohort studies. These would be expected to be less precise than randomised controlled trials but both types of study have advantages and disadvantages. In particular, there are inherent difficulties in designing an epidemiological randomised controlled trial to investigate the effects of sea bathing that reduce the precision relative to a conventional randomised controlled

drug trial. WHO (1998) focuses on Kay *et al.*'s (1994) dose-response model which links gastroenteritis to faecal streptococci counts measured at the location and time of exposure and is based on data from the UK randomised controlled trials. Superficially this is a logical choice but the decision to disregard the cohort studies is questionable, particularly given serious weaknesses in the design, conduct and analysis of the UK randomised controlled trials, which will be described in Section 3.

3 The UK randomised controlled trials

3.1 OVERVIEW

A series of four randomised controlled trials and a parallel series of ten prospective cohort studies were conducted in the UK during the period 1989-1992. Both series were jointly funded by the Department of the Environment (DoE), the Department of Health, the Welsh Office, and the National Rivers Authority. The main contractor for the studies was the Water Research Centre plc (WRc). The randomised controlled trials were sub-contracted to the Centre for Research into Environment and Health, St David's University College, Lampeter. The cohort studies were sub-contracted to the Institute of Public Health, University of Surrey.

This section provides a detailed description and assessment of different aspects of the randomised controlled trials that are reported in Kay *et al.* (1994). Reference is made to information that has appeared in various documents up to and including Kay *et al.* (1994). A chronological list of the documents is given below. In many of the documents the randomised controlled trials and the prospective cohort studies are referred to as (controlled) cohort studies and beach surveys, respectively.

- Pike (1991a) is the final report to the DoE of Phase I of the randomised controlled trials and cohort studies. It describes pilot studies of each type conducted at Langland Bay in August and September 1989. Final reports to the DoE by each of the sub-contractors are presented as appendices to the main report.
- Pike (1991b) is the final report to the DoE of Phase II of the randomised controlled trials and cohort studies. It describes a randomised controlled trial conducted at Moreton and a prospective cohort study conducted at Ramsgate in August 1990. Final reports to the DoE by each of the sub-contractors are presented as appendices to the main report.
- Jones *et al.* (1993) is the final report to the DoE of Phase III of the randomised controlled trials. It describes trials conducted at Southsea in July 1991 and at Southend-on-Sea in July 1992.

- Balarajan (1993) is the final report to the DoE of Phase III of the prospective cohort studies. It describes studies conducted at Lyme Regis, Morecambe, Paignton, and Rhyl in August 1991, and at Cleethorpes, Instow, Skegness, and Westward Ho! in August 1992.
- Pike (1994) is the final report to the DoE of Phase III of the randomised controlled trials and cohort studies. Each series of studies is discussed in its entirety and the results of the two types of studies are compared. It should be noted that the review by WHO (1998) includes the results of the cohort studies presented in this report but ignores those of the randomised controlled trials.
- Kay *et al.* (1994) is a peer-reviewed article that presents results of the randomised controlled trials relating to one type of health outcome (gastroenteritis).

Additional documents based on the UK randomised controlled trials have appeared since the publication of Kay *et al.* (1994). These will be discussed in Section 4 when the results reported by Kay *et al.* (1994) are compared with those reported in other articles.

3.2 DESIGN

Each of the four randomised controlled trials was conducted according to the same basic design. Subjects were recruited from the local population in the weeks leading up to the trial. Pike (1991a, b), Jones *et al.* (1993), Pike (1994) and Kay *et al.* (1994) do not state whether the trial subjects were recruited from the local resident population, the tourist population or a mixture of the two. It would be useful to have this information.

Each trial took place on a single afternoon during the normal bathing season. For ethical reasons, subjects had to be more than 18 years old, and the beach used for the trial had to have met current EC mandatory standards in the previous bathing season. On the trial day (exposure day), subjects arriving at the beach were randomised to

bathing or non-bathing groups. Those who bathed were asked to spend at least 10 minutes in the water and to immerse their heads at least three times.

Water samples were collected from the bathing area at half-hourly intervals during the exposure period, at 20 m intervals along the strip of beach designated for bathing, and at surf, mid, and chest depths. The samples were analysed for bacterial indicators of water quality. Water quality measurements were subsequently assigned to each bather, according to the time, location, and depth at which they bathed. Further details relating to exposure assessment are given in Section 3.3.

Subjects completed four interviews during the course of the trial. The first (pre-exposure) interview took place not more than two days before exposure day. Subjects were asked to provide information about themselves, including age, sex, address where they could be contacted for follow-up interviews, general state of health, recreational activities involving contact with water, and other potential confounding factors associated with anticipated adverse health outcomes. Subjects also underwent a medical examination. Those who were regarded as unfit to participate (because of serious heart conditions, life-threatening illnesses, etc.) were excluded from the rest of the trial. The second (exposure-day) interview took place at the beach. Subjects were asked to provide information about their health, food intake, leisure activities, and other potential confounding factors in the days leading up to the trial. In the third (one-week post-exposure) interview subjects were asked to provide information about their health, food intake, leisure activities, and other potential confounding factors in the week following exposure. Subjects also underwent a medical examination which included collection of faecal specimens and ear and throat swabs. The fourth (three-weeks post-exposure) follow up was conducted by postal questionnaire. Subjects were asked about their health, food intake, leisure activities, and other potential confounding factors in the two weeks since the previous interview. They were also asked to provide faecal specimens.

Further details on the design of the randomised controlled trials, together with comments on the statistical analysis and interpretation of results reported by Kay *et al.* (1994) are presented in the remainder of Section 3.

3.3 EXPOSURES

The randomised controlled trials were designed to study health effects in relation to the quality of the actual water to which bathers were exposed. This was achieved by the exposed subjects bathing in a defined area of water which was sampled very frequently throughout the study period. WHO (1998) highlights the need for accurate assessment of ingestion of water and the degree and duration of water contact. These issues were adequately addressed by the randomised controlled trials (Pike, 1994). Exposed subjects bathed in 20 m wide roped-off strips; in 1989 and 1990 there were five strips over a 100 m stretch of coastline, whereas in 1991 and 1992 there were three strips over 60 m of coastline. Bathers entered the water in one of the roped-off 20 m subsections and were instructed to immerse their heads fully three times and stay in the water for at least ten minutes. On leaving the water, bathers were asked whether or not they had swallowed water.

Water was sampled every 30 minutes (every 20 minutes in 1989 and 1990) over the exposure period in every 20 m interval and at surf depth, mid depth (30 cm below the surface in 1 m depth of water - as specified by the EC Bathing Water Directive) and at chest depth (30 cm below the surface in 1.3-1.4 m depth of water). Chest depth was the location at which immersion took place (Kay *et al.*, 1994). Water quality measurements were assigned to bathers according to the location and time of exposure. Measurements at chest depth do not appear to have been made in previous studies, mid depth being the most common. Although Kay *et al.* (1994) do not explicitly comment on this, chest depth appears to be the most appropriate depth at which to assess personal exposure because it was the location at which immersion took place.

Water samples were transferred to a mobile laboratory for analysis within six hours. Environmental conditions at the time of sampling were also noted. Water samples were analysed by membrane filtration and subsequent incubation on selective media. Indicator organisms measured in the samples were:

- total coliforms;
- faecal coliforms;
- faecal streptococci;
- total staphylococci; and
- *P. aeruginosa.*

Duplicate water samples were collected for quality control purposes and examined at the same time as the main samples. The same sub-contractor was used to analyse the water samples in all four trials, thus avoiding inter-laboratory variability. The sub-contractor had received accreditation for its microbiology laboratory from the National Accreditation of Measurement and Sampling (NAMAS) prior to the 1990 trial. The approaches to analytical quality control differed in the four trials, ranging from seven duplicate samples in 1992 to 18 in 1989 (Pike, 1994).

3.4 HEALTH EFFECTS

Data on the following classes of self-reported symptoms were collected during interviews and via the postal questionnaires:

- flu/cold;
- chest;
- ear/eye;
- gut;
- skin; and
- other (including excessive tiredness, dizziness, pins and needles, and muscle cramps).

The questionnaires asked those who indicated any illness if they had obtained medication, or visited a doctor, or missed any days of work/normal activity or if they had been hospitalised. The questionnaires also asked for details regarding the onset and duration of symptoms; this was repeated in the one- and three-week post-exposure interviews, which enabled cross-checking between the two to be carried out.

Occurrence of gastrointestinal illness was chosen for detailed analysis because it had been most often examined in international epidemiological studies seeking to inform policy on recreational water standards (Jones *et al.*, 1993). Kay *et al.* (1994) report results for objective gastroenteritis (see Section 2.3).

3.5 CONFOUNDERS

Potential confounding factors on which data were collected as part of the randomised controlled trials are listed in Annex I. It is impossible to say exactly which of these factors do actually confound the relationship between gastroenteritis and exposure because of the limited and largely inappropriate way in which their effects were investigated (see Section 3.7). The only factors considered to be potential confounders by Kay *et al.* (1994) are those listed in their Table 1 (see Annex I). Of these, the most important are probably:

- age;
- sex;
- frequency of diarrhoea;
- illness in the four weeks preceding exposure day that lasted for more than 24 hours;
- consumption of certain foods associated with gastroenteritis;
- illnesses in the household in the three weeks following exposure; and
- additional bathing in the period from three days before exposure to three weeks afterwards.

Duration of exposure and ingestion of sea water were analysed by Jones *et al.* (1993) and Pike (1994) but not as potential confounders (see Section 3.7) and the results were not presented by Kay *et al.* (1994). The effects of the remaining factors listed in Annex I were apparently never analysed. Of these, the following might be expected to influence risks of experiencing or reporting gastroenteritis:

- awareness of beach maintenance in the UK;
- awareness of pollution;
- awareness of news/media coverage of the trials; and
- holidays/business trips in the UK and abroad in the period from four weeks before exposure to three weeks afterwards.

Ideally, statistical analysis of relationships between gastroenteritis and water quality indicators should incorporate adjustment for all possible confounders and all other potential sources of variability in risks of experiencing gastroenteritis (for example, geographical location). At the very least, all confounders and other variables that are significantly associated with gastroenteritis at, say, the 25% level (p=0.25) should be included in models attempting to relate gastroenteritis to water quality, and interactions between confounders and water quality indicators should be considered. Failure to do so might result in seriously misleading estimates of risk at any given level of water quality but particularly for mildly polluted water (see Section 3.7).

3.6 SOURCES OF BIAS

Overview

Several sources of bias are apparent from Kay *et al.*'s (1994) description of the randomised controlled trials. Two other important sources of bias are not (entirely) evident from Kay *et al.*'s (1994) descriptions of the trials but can be deduced from earlier reports. The sources of bias and their expected effects on estimates of risks of experiencing gastroenteritis are summarised in Table 3.1.

Table 3.1 Sources of bias in the UK randomised controlled trials and their expected effects on estimated risks of experiencing gastroenteritis

Source of bias	Estimate		
	Baseline risk (non-bathers)	**Risk to bathers**	**Relative risk (bathers:non-bathers)**
Restriction to healthy adult volunteers	↓	↓	↓↑
Self-reporting of symptoms	↑	↑	↑
Length of recall period	↓	↑	↑
Definition of gastroenteritis	↑	↑	Not affected
Length of observation period	↑	↑	↓
Repeated statistical testing	Not applicable	Not applicable	↑
Model-selection bias	Not applicable	Not applicable	↑
Recruitment methods	↓	↑	↑
Timing of exposure-day interviews	↓	↑	↑

↓ Indicates a tendency towards under-estimation ↑ Indicates a tendency towards over-estimation

Restriction to healthy adult volunteers

The exclusion of children and sick adults, together with an apparent under-representation of elderly people (the median age of the volunteers was approximately 32) suggests that the population of volunteers contained fewer individuals with a potentially weakened immune system or susceptibility to gastrointestinal problems than the general population. This would be expected to lead to under-estimation of both the baseline risk of gastroenteritis (that is, the risk to non-bathers) and the risk to bathers. Kay *et al.* (1994) and WHO (1998) both recognised this source of bias.

Self-reporting of symptoms by subjects

Although subjects underwent medical examinations at various stages of the trials, most of the information about illnesses, and certainly that used in the analyses reported by Kay *et al.* (1994), was obtained by interview or postal questionnaire. The tendency for subjects to report even very mild symptoms would be expected to lead to over-estimation of both the baseline risk and the risks to bathers.

Length of recall required

Subjects were asked to provide information about illnesses and other factors retrospectively. The data collected in the initial and three-week post-exposure interviews would have been most seriously affected since these interviews required the longest periods of recall (up to four and three weeks, respectively). Differential recall between bathers and non-bathers could lead to over-estimation of the relative risk of bathing (bathers compared with non-bathers).

Definition of gastroenteritis

Kay *et al.*'s (1994) definition of objective gastroenteritis is slightly less stringent than the definition of highly credible gastroenteritis used by Cabelli *et al.* (1982) to approximate clinically defined viral gastroenteritis. This source of bias would be expected to lead to over-estimation of both the baseline risk of gastroenteritis and the risk to bathers.

Three-week observation period for gastrointestinal symptoms

Kay *et al.* (1994) base their analyses on reports of gastroenteritis up to three weeks after exposure. Interpretation of their results needs to take into account the plausibility of cases of gastroenteritis that arose towards the end of the study period being due to bathing on exposure day (see Section 2.3). Only 34.3% of all the cases of gastroenteritis developed within one week of exposure (Kay *et al.*, 1994), which suggests that this source of bias might lead to substantial over-estimation of both the baseline risk of gastroenteritis and the risk to bathers.

Repeated statistical testing

Kay *et al.* (1994) analysed the data on gastrointestinal symptoms in at least 50 different ways: three separate tests were used to examine dose-response models

relating gastrointestinal symptoms to water quality for each of five indicator bacteria measured at each of three depths; for the one indicator/depth combination that was found to have a statistically significant relationship with gastroenteritis (faecal streptococci measured at chest depth), five further tests were conducted - see Section 3.7. Statistical theory states that 5% of tests (one in 20) will yield a statistically significant result at the normally quoted level of significance (5%) purely by chance under the null hypothesis. We would, therefore, expect at least two of the 50 tests conducted by Kay *et al.* (1994) to have given a statistically significant result just by chance even if there were no real effect. If all the statistically significant results related to positive correlations with gastroenteritis, then this could lead to over-estimation of the number of significant relationships.

Model-selection bias

Kay *et al.* (1994) chose to model the relationship between gastroenteritis and faecal streptococci measured at chest depth because this combination of indicator bacteria and depth gave the most significant relationship. The model that fits a given set of data most closely will necessarily have smaller standard errors (and confidence intervals) than any other model considered. When Kay *et al.*'s (1994) final dose-response model is used to predict risks of gastroenteritis from particular concentrations of faecal streptococci the standard errors (and confidence intervals) attached to predictions should, therefore, be increased (widened) to take this bias into account. This form of bias could lead to over-estimation of the relative risks of bathing (bathers compared with non-bathers).

Recruitment methods

Subjects were recruited from local population centres by means of publicity which included press coverage, media broadcasts, subject information sheets and, in the case of the later trials, displays showing information from the previous trials (Pike, 1991a, b; Jones *et al.*, 1993). The subject information sheets identified stomach infections as

a potential health outcome. This form of bias could lead to over-estimation of the risks of bathing by over-alerting bathers to the sort of health problems that might occur following exposure.

Timing of exposure-day interviews

The trial methodology described in Kay *et al.* (1994) suggests that some of the exposure-day interviews might actually have taken place after the exposure period rather than before as the protocol required. Knowledge of bather/non-bather status during exposure-day interviews coupled with knowledge gained from the subject information sheets of the fact that bathing carries a risk of stomach infections might have led to over-reporting of symptoms associated with gastroenteritis by bathers relative to non-bathers in the exposure-day interviews. In the Southend-on-Sea trial, for example, relative risks of loose motions and diarrhoea on exposure day were 2.81 and 9.27 ($p=0.023$ and 0.021), respectively (Jones *et al.*, 1993). The possibility of bias arising from timing of the exposure-day interviews is confirmed by information in earlier reports of the randomised controlled trials. On arrival at the beach, subjects were given two lists of names, one printed in blue, the other in red. The lists contained names of pre-registered subjects that had been randomised to either the bathing group (blue) or the non-bathing group (red), together with identification numbers of supervisors to whom subjects were asked to report (Pike, 1991a, b; Jones *et al.*, 1993). Subjects completed their exposure-day interviews after receiving the blue and red lists. Even if the interviews took place before exposure itself, at a stage when subjects had not been told that blue names were for bathers and red names were for non-bathers, it would not have been difficult to guess the significance of the colour coding. Indeed, some subjects were recruited at the beach on exposure day itself, and their initial and exposure-day interviews could both have taken place after the exposure period. Over-reporting of symptoms by bathers as a result of the timing of exposure-day interviews could lead to over-estimation of risks of bathing.

Summary

While it is not certain that the effects of bias would be as shown in Table 3.1, it is to be expected that such effects did occur. The main conclusion to be drawn from Table 3.1 is that risks to bathers and relative risks (bathers compared with non-bathers) could have been seriously over-estimated. Furthermore, it is likely that risks to non-bathers would have been over-estimated. The only source of bias that was strongly highlighted by Kay *et al.* (1994) and WHO (1998) was the one source for which under-estimation of risks to bathers would be expected to occur (restriction to healthy adults). WHO (1998) also acknowledges the potential for media and publicity to induce bias in randomised controlled trials but fails to mention the extent to which this form of bias could have affected the trials reported by Kay *et al.* (1994).

Some sources of bias listed in Table 3.1 arise because of inherent difficulties in designing randomised controlled trials to investigate the health effects of sea bathing (inability to blind subjects to exposure status, ethical considerations, etc.). Other forms of bias appear to have been introduced because the investigators did not adhere to the study protocol as rigidly as they might have done. Neither of these types of bias can be eliminated retrospectively but interpretation of the study results should attempt to take them into account. The remaining sources of bias (definition of gastroenteritis, three-week observation period for gastrointestinal symptoms, repeated statistical testing and model-selection bias) arose during statistical analysis and could be addressed by appropriate re-analysis of the data (see Section 3.8).

3.7 STATISTICAL ANALYSIS, RESULTS AND INTERPRETATION

Overview

The main elements of Kay *et al.*'s (1994) analysis of the relationship between gastrointestinal symptoms and exposure were as follows:

- Subjects with incomplete data were excluded.

- The effectiveness of randomisation was investigated.
- A subset of significant confounders to be controlled for in later analyses was selected.
- Linear trend in the risks of gastroenteritis in various categories of exposure (unexposed, exposed to water of good quality, exposed to water of poor quality) was analysed for each of five indicator bacteria at each of three depths (surf, mid, and chest) at each of the four trial sites in turn.
- Linear trend in the risks of gastroenteritis was re-analysed excluding non-bathers.
- Differences in trends between the four trials were analysed.
- Indicator bacteria and depth combinations with significant trends were selected for further analysis. In the event, only faecal streptococci at chest depth was significant and no differences between trial sites were detected. All further analyses were conducted using faecal streptococci measured at chest depth after pooling data from the four trials.
- Linear trend was re-analysed using quantiles of faecal streptococci distributions to define categories.
- Linear trend was re-analysed using 20-unit intervals of faecal streptococci concentrations to define categories.
- The dose-response relationship between gastroenteritis and faecal streptococci concentrations was analysed using a further categorisation method and allowing for significant confounders.
- The dose-response relationship between gastroenteritis and faecal streptococci concentrations was analysed excluding non-bathers, treating faecal streptococci concentration as a continuous variable, and ignoring confounders.

Each of these elements is discussed in detail in below.

Exclusion of subjects

Potential subjects were excluded from the entire study if initial interviews revealed them to be medically unfit to take part or afraid of water. Subjects who failed to turn up on the trial day, or failed to complete either of the post-exposure interviews, or who failed to comply with their randomisation status, or for whom precise details of water quality could not be determined were excluded from the statistical analysis. Crude rates of gastrointestinal symptoms reported on the day were much higher amongst bathers (5.13%) than non-bathers (0.53%) (p=0.021; Jones *et al.*, 1993). Subjects who reported gastrointestinal symptoms on the day were, therefore, excluded from the analysis (Pike, 1994) although Kay *et al.* (1994) does not mention this. The total number of subjects included in the statistical analysis was 1112; of these, 507 were bathers and 605 were non-bathers (Pike, 1994).

Effectiveness of randomisation

The successfulness of randomisation was assessed by comparing distributions of non-water-related risk factors for gastroenteritis among bathers and non-bathers. Only those potential confounders identified in Kay *et al.* (1994, Table 1) were included (see Annex I). Each confounder was assessed in turn using univariate chi-squared tests. Kay *et al.* (1994) argue that where randomisation was successful, crude rates of gastroenteritis for bathers and non-bathers would contain an implicit adjustment for the effects of the confounders concerned. This is true to the extent that equal representation of different levels of confounders in the bathing and non-bathing groups would implicitly correct for the effects of those confounders. Analysis based on the crude rates of gastroenteritis would give good estimates of the relative risks of bathing in the study population. Partial extrapolation of the results to the general population would, however, be facilitated by including all possible confounders in the analysis explicitly. This would, for example, allow the amount of variation in risk due to particular confounders to be examined as part of risk assessment (see Section 5). Where randomisation was not successful, Kay *et al.* (1994) assessed the effects of the

corresponding confounders to see whether they changed the relationship between gastrointestinal symptoms and water quality (see below).

Analysis of linear trend

Linear trends between gastroenteritis and exposure (measured using five bacterial indicators at each of three depths) were assessed using Mantel-Haenszel chi-squared tests. Where possible, exposure was categorised according to whether water quality was below or above existing standards. The standards used were: NTAC (1968) for total coliforms; EPA (1976) for faecal coliforms; EPA (1986) and MNHW (1992) for faecal streptococci. For the other indicators no standards exist, and so exposure was categorised according to whether water quality was below or above the median of the observed water quality measurements. Non-bathers formed a further category. For any given bacterial indicator, these analyses yielded rates of gastroenteritis in each category of exposure. It should be noted that the categories of exposure for all indicators except faecal streptococci were dichotomous. Table 3.2, column 1 shows the categorisation used for faecal streptococci. Thirty-five faecal streptococci per 100 ml is the geometric mean criterion of the EPA (1986) and Canadian (MNHW, 1992) guidelines for marine recreational waters (Pike, 1994). Neither Pike (1994) nor Kay *et al.* (1994) offer any explanation as to why faecal streptococci counts of 35 per 100 ml and above were subdivided into two categories. Moreover, given the level of detail in the measurements of the indicators, they could all have been analysed as continuous variables.

Table 3.2 Basis of categorisation of faecal streptococci concentrations used in Kay *et al.*'s (1994) statistical analysis; units are counts per 100 ml

Existing standards	Quantiles	Twenty-unit intervals	Bathers only
Unexposed	Unexposed	Unexposed	
Exposed 0-34	Exposed 0-13	Exposed 0-19	Exposed 0-39
Exposed 35-69	Exposed 14-26	Exposed 20-39	Exposed 40-59
Exposed 70 +	Exposed 27-49	Exposed 40-59	Exposed 60-79
	Exposed 50-158	Exposed 60-79	Exposed 80 +
		Exposed 80 +	

The trend analyses were repeated excluding the unexposed group. Kay *et al*. (1994) interpreted the significance levels of these tests to take account of repeated testing: for any given bacterial indicator and depth combination, a linear trend was deemed to be significant only if the Mantel-Haenszel chi-squared statistic yielded $p<0.05$ for both tests (that is, with and without the unexposed group).

Trends were further assessed by study site. In all the tests described above, the only indicator and depth combination which yielded a significant linear trend was faecal streptococci measured at chest depth. No significant differences between trial sites were detected, and so pooled data for the four trials were used in the subsequent analyses. It is reassuring that, despite the repeated statistical testing, faecal streptococci measured at chest depth (the depth at which exposure actually occurred) was found to have the strongest relationship with gastroenteritis.

Linear trends in gastroenteritis symptoms in response to faecal streptococci concentrations were further analysed using different categorisation schemes (quantiles, and 20-unit intervals - see Table 3.2, columns 2 and 3, respectively). Kay *et al.* (1994) argue that these tests were used to examine the dose-response relationship between gastroenteritis and water quality in ever-increasing detail. With this aim in mind, it would have been preferable to have treated faecal streptococci concentration as a continuous variable (see below). Also, it is appropriate to choose a categorisation method *a priori* and to try other categorisations if a significant effect is found as a check on whether the significant result is an artefact of the initial categorisation. It is inappropriate to try many different categorisations in order to seek out the one which gives the most significant effect (Rothman & Greenland, 1998).

Dose-response modelling treating water quality as a categorical variable

Having established a linear trend in the rates of gastroenteritis in response to increasing concentrations of faecal streptococci measured at chest depth, Kay *et al.*

(1994) carried out a further analysis which was restricted to bathers and used a further categorisation of faecal streptococci concentrations (see Table 3.2, column 4). In this analysis, multiple logistic regression was used to examine the effects of confounding factors found to be significant in the earlier chi-squared tests. These factors were: age; sex; gastroenteritis in the household that preceded the bather's own symptoms; and a composite variable representing other non-water-related risk factors (Kay *et al.*, 1994, Table 6). The composite variable combined data on: frequency of diarrhoea; diarrhoea lasting more than 24 hours in the three weeks preceding exposure day; unusual fatigue lasting more than 23 hours in the three weeks preceding exposure day; and consumption of hamburgers, purchased sandwiches, or take-away food during the period from three days before exposure day to seven days afterwards. The exact method of calculating the composite variable was not reported. It is unclear whether, for example, predisposition to diarrhoea plus consumption of hamburgers counts differently from one of these factors alone.

Kay *et al.* (1994) criticise other studies for not measuring or adequately controlling for non-water-related risk factors. In their own study, however, they failed to investigate all potential confounding factors in the chi-squared tests described earlier. They also failed to allow for confounding factors that were not significant individually but might have had an appreciable effect on the dose-response relationship in combination with other confounders. It would have been preferable to have included all potential confounders in the logistic regression model. This would have allowed the dose-response relationship between gastroenteritis and faecal streptococci to have been estimated after adjusting for all additional sources of variation (beyond water quality). It is impossible to say exactly how the analysis might have been changed had all confounders been included. Of the confounders that were omitted, one might expect the following to have considerable impact on the risks of gastroenteritis: additional bathing in the period from three days before exposure to three weeks afterwards; holidays/business trips in the UK and abroad in the period from four weeks before exposure to three weeks afterwards; awareness of beach maintenance and pollution in

the UK; news/media coverage of the trials; duration of exposure; and ingestion of water. It would also have been possible to have included terms in the regression model to account for differences between trial sites. Even though differences in trends between sites were not significant, inclusion of such terms would have allowed this source of variability to be quantified. Quantification of the variability in risks due to differences between trial sites would have been useful in risk assessment (see Section 5).

Kay *et al.* (1994) argue that 'Logistic regression analysis ... revealed no confounding of the relation between exposure to faecal streptococci and illness' by the risk factors shown in their Table 6. The table contains odds-ratios for the different levels of exposure and the confounding factors from the model that includes the confounders; equivalent results for a model excluding the confounders would be needed in order to draw conclusions about the extent of confounding. Table 6 of Kay *et al.* (1994) shows that, after adjustment for the (limited) set of confounders, increased faecal streptococci concentrations were associated with statistically significant increases in the risks of gastroenteritis. It should be noted, however, that the highest odds-ratio in Kay *et al.* (1994), Table 6 is not for the effect of faecal pollution but for the effect of gastroenteritis in the household that preceded the subject's own symptoms.

Logistic regression excluding non-bathers has advantages and disadvantages. The main advantage is that bathers were blind to the exact quality of the water they were bathing in, and so estimates of the relative risks of bathing in mildly and more heavily polluted waters should be fairly reliable. WHO (1998) highlights the fact that comparison of bathers as a whole with non-bathers on the beach may lead to under-estimation of the effects of water quality (because viruses may be transferred from the water to the air) and recommends that dose-response relationships between bathers in mildly polluted waters and those in more heavily polluted waters should be used whenever possible. One disadvantage of excluding non-bathers in Kay *et al.*'s (1994) analysis is that the sample size was reduced from 1112 to 507. The larger sample size

that would have resulted from inclusion of non-bathers would have yielded more precise estimates of the risks of gastroenteritis in relation to water quality (smaller standard errors and narrower confidence intervals). Another disadvantage is that the effects of sea water itself cannot be investigated. Fitting a model to data for non-bathers as well as bathers, and inclusion in the model of an indicator variable representing exposure (bather/non-bather) status would have allowed the effects of contact with sea water to be examined separately from the effects of water quality (see Section 6). Note that this would not necessarily lead to a good estimate of the effect of clean sea water because the coefficient for the indicator variable would be confounded with the effects of all the characteristics in which bathers differed from non-bathers.

Dose-response modelling treating water quality as a continuous variable

Kay *et al.* (1994) fit one further dose-response model to the relationship between gastroenteritis and faecal streptococci concentrations. This is the model which WHO (1998) uses as the basis for its risk assessment exercise. The model is restricted to bathers, and although Kay *et al.* (1994) imply that the model adjusts for potential confounding factors, the only term in the model is that for faecal streptococci concentrations (Jones *et al.*, 1993). This is, in fact, evident from Kay *et al.* (1994): if the model explicitly adjusted for confounders then an explanation would be needed of the levels of the confounding factors to which the graph of the fitted relationship corresponded (for example, men or women, which ten-year age group, etc.).

In the final analysis faecal streptococci is treated as a continuous variable. The fitted model allows estimates of the risk of gastroenteritis to be calculated for any specified faecal streptococci count. All the other models yielded estimates of risk for the category of exposure in which the specified faecal streptococci count fell. A preferred approach might have been to treat water quality measurements as continuous variables throughout the analysis if the desired product of the analysis was detailed expression

of the dose-response relationship. Kay *et al.* (1994) argue that square roots of the faecal streptococci counts should be used in fitting the model. The statistical distribution of faecal streptococci counts is immaterial when the counts are used as an explanatory variable in a regression model. The most appropriate transformation of the counts should be determined by examining the fit of the relationship with the log-odds of experiencing gastroenteritis. It would be convenient to have a regression model in which the $\log_{10}$ transformation of the faecal strepococci counts is used as an explanatory variable because the distribution used in the WHO risk assessment is of this form (see Section 5). However, the $\log_{10}$ transformation is not necessary and it may not even be appropriate.

The final model actually consists of two separate logistic models, one fitted to data on bathers who experienced less than 32 faecal streptococci per 100 ml, and the other fitted to data on bathers who experienced at least 32 faecal streptococci per 100 ml. This cut-off was chosen as the median value of the category of exposure (20-39 faecal streptococci per 100 ml) that had indicated no excess risk among bathers in any of the previous analyses. The logistic regression model fitted to data on bathers who experienced less than 32 faecal streptococci per 100 ml showed no significant increase in risk with decreasing water quality. Kay *et al.* (1994) suggest that this provides evidence for a threshold of risk at 32 faecal streptococci per 100 ml. This is not a valid statistical method for determining a threshold but rather a method of trial and error. In any case, a statistically significant increase in risk at 32 faecal streptococci per 100 ml does not necessary imply a sufficiently important risk to warrant intervention on the grounds of public health (this would depend on the clinical significance of the increase in risk). Risk assessment can, however, be used to combine the entire dose-response relationship and the likely distribution of faecal streptococci concentrations to yield guideline values (see Section 5). It should be noted that, despite the lack of statistical validity for selection of the threshold of 32 faecal streptococci per 100 ml, this figure reflects the level of water quality at which statistically significant increases in risk were observed in the earlier analyses based on categorical expression of water quality.

For example, the category 40-59 faecal streptococci per 100 ml had an odds ratio of 1.91 (95% confidence interval 1.60-2.28) relative to the baseline category of 0-39 faecal streptococci per 100 ml (Kay *et al.,* 1994, Table 6). However, uncertainties as to the reliability of the earlier analyses because of potential sources of bias and improper control for confounders render this threshold misleading. Further, it seems unlikely that there is zero risk at concentrations below 32 faecal streptococci per 100 ml, and non-zero risk thereafter. A more plausible model would be for a continuously increasing risk, even at very low faecal streptococci concentrations. Such a model could be fitted to the data for both bathers and non-bathers using logistic regression analysis.

Kay *et al.* (1994) express the fitted model in terms of 'excess' risk attributable to increasing faecal streptococci concentrations. The logistic model implies a multiplicative increase in risk (that is, the risk of gastroenteritis at a specified faecal streptococci concentration is estimated to be some multiple of the risk at a baseline level of exposure). Expression of the model in terms of excess risk implies that risks due to different sources are additive. Further problems associated with the representation in terms of excess risk are that it is necessary to specify what the baseline risk corresponds to (for example, bathers in mildly polluted waters, or non-bathers), and that the estimates of the baseline risks must be reliable. Kay *et al.*'s (1994) model suffers from potential bias in the estimates of the baseline risk (see Section 3.6).

A further problem with Kay *et al.'s* (1994) final dose-response model is that neither standard errors nor confidence intervals for estimated risks are reported. Such information would allow assessment of uncertainty in the fitted model that could be attributed to using a finite sample of subjects rather than the entire population of interest. The smaller the sample size, the larger (wider) standard errors (confidence intervals) would be expected to be. Alternatively, the goodness-of-fit of the statistical model to the observed data could be discussed.

Related analyses

WHO (1998) argues that concomitant outcomes should be measured, analysed and reported. Kay *et al.* (1994) do not mention that data from medical samples and information about days off work due to illness and visits to family doctors failed to confirm the risks of gastroenteritis due to bathing (Pike, 1994). These results might have arisen because gastroenteritis is a relatively minor illness, and pathogens that might have been responsible for some of the symptoms were not detectable by the methods used (Pike, 1994).

An aspect of the UK randomised controlled trials that was overlooked by Kay *et al.* (1994) but reported by Jones *et al.* (1993) and Pike (1994) is the relationship between gastrointestinal symptoms and exposure variables other than water quality (ingestion and duration of exposure). There was a significant correlation between gastrointestinal symptoms and ingestion of water in certain trials (Pike, 1994). The average time spent in the water varied between the different trials (being greatest at Langland Bay and Moreton), but duration of exposure did not differ between individuals who experienced gastrointestinal symptoms and those who did not (Jones *et al.*, 1993). Unfortunately, the analysis of the relationship between duration of exposure and gastrointestinal symptoms was invalid: Jones *et al.* (1993) treated duration of exposure as the outcome (response) variable and occurrence of gastrointestinal symptoms as a potential explanatory variable. Ingestion of water and duration of exposure should have been included as potential confounders in the logistic regression models relating gastrointestinal symptoms to exposure.

Kay *et al.* (1994) also failed to mention that a statistically significant difference was found between the seven quality control duplicate samples for faecal streptococci in the 1992 trial (Jones *et al.*, 1993). For the purpose of measuring bacterial indicator counts it would be good practice to collect triplicate water samples at each time, location and depth, and to use the geometric mean of the resulting counts in

subsequent statistical models. This would, however, give rise to considerable practical and logistical difficulties.

3.8 SUMMARY

The UK randomised controlled trials in general, and the results reported by Kay *et al.* (1994) in particular, display several positive features. There are, however, many serious weaknesses which limit the extent to which Kay *et al.*'s (1994) final dose-response model can be regarded as valid in relation to the UK randomised controlled trials themselves and reliable in the wider context of microbiological standard setting for the protection of human health.

The main strengths of Kay *et al.*'s (1994) dose-response model are that it links incidence of gastroenteritis to faecal streptococci concentrations (these being the most commonly reported adverse health effect in other studies and the bacterial indicator most frequently associated with it), and that it is based on a series of randomised controlled trials in which precise measurements of exposure (degree and duration of water contact and whether or not ingestion of water occurred) were collected.

One of the major drawbacks with Kay *et al.*'s (1994) dose-response model is the extent to which risks of experiencing gastroenteritis are expected to have been over-estimated because of bias in the design, conduct and statistical analysis of the UK randomised controlled trials. Risks to non-bathers, risks to bathers and relative risks (bathers compared with non-bathers) are all expected to have been over-estimated. Concerns about Kay *et al.*'s (1994) dose-response model are heightened by inadequacies in the statistical analysis. The major inadequacies are:

- insufficient controlling for confounders;
- failure to treat water quality measurements as continuous random variables from the outset;

- insufficient consideration of the most appropriate transformation of faecal streptococci counts to use in dose-response modelling;
- imposition of a subjective threshold of risk in dose-response modelling for which no justification was offered on the grounds of biological plausibility;
- reduction in sample size by the exclusion of non-bathers;
- conflict between the multiplicative nature of the logistic regression model actually fitted and the additive form of model that would be required in order to express the fitted model in terms of excess risk;
- failure to report standard errors and confidence intervals for the fitted dose-response model; and
- failure to discuss the results of the medical examinations.

It is recognised that space considerations play a part in the preparation of manuscripts for publication in peer-reviewed journals. Given the large number of details that were omitted from Kay *et al.* (1994) and some fundamental weaknesses in the statistical analysis it is, however, surprising that the article was deemed to be acceptable for publication in the form in which it appeared. All the short-comings in the statistical analysis and several sources of bias could be eliminated by appropriate re-analysis (see Section 6). The extent to which the WHO guideline document is justified in relying on Kay *et al.*'s (1994) dose-response model is discussed in Section 4.

4 The UK randomised controlled trials in context with other studies

4.1 REVIEW OF OTHER STUDIES

Table 4.1 gives a brief description of the studies reviewed in WHO (1998, Chapter 4) together with a number of additional studies published since 1997. Details are given of study dates, locations, designs, sample sizes and what was measured in relation to health outcomes and indicator organisms. Significant aspects of the study are commented on, including how exposure was defined for bathers and whether or not there were specific criteria to be fulfilled, such as full head immersion. The studies are divided into three categories according to the frequency of water quality measurements: those using seasonal averages only; those collecting daily measurements but using study-period averages; and those using daily measurements. This review is concerned with studies into health risks associated with sea bathing, and so only those studies conducted in marine waters are listed in Table 4.1, whereas WHO (1998, Table 4.1) includes results for both marine and freshwater studies. Indicator organisms and pathogens have different survival rates in fresh and marine waters, and so consideration of risks from the different water types should be kept completely separate.

It should be noted that several of the studies listed in Table 4.1 gave rise to more than one published report. This is particularly true of the UK epidemiological studies: Fleisher *et al.* (1996) is a peer-reviewed article that presents results from the UK randomised controlled trials for non-enteric outcomes (acute febrile respiratory illness, and eye, ear and skin complaints); and van Dijk *et al.* (1996) is the final report to the DoE of a contract awarded to the WRc in 1995 to undertake further statistical analysis of the UK cohort studies.

Table 4.2 summarises the results of those studies listed in Table 4.1 which specifically relate the incidence rate of gastroenteritis to water quality measurements using faecal streptococci, enterococci, *E. coli* and/or faecal coliforms. Relative risks and/or incidence rates for gastroenteritis are given for different exposure categories for the most relevant indicator organisms. The comment column gives information as to whether or not any account was taken of potential confounders such as age or recent food intake. Table B.1 in Annex II gives the corresponding information for studies relating gastrointestinal symptoms to other bacterial indicators. Table B.2 in Annex II summarises the results of studies relating other outcomes to water quality indicators.

A number of discrepancies exist between the ways in which results are presented in different articles. Some quote relative risks for bathers as compared with non-bathers (for example, Bandaranayake *et al.*, 1995). Many others quote relative risks compared with bathers in the lowest exposure category (for example, Haile *et al.*, 1999). The latter is used in WHO (1998, Table 4.4), although the table legend states that the comparison is between bathers and non-bathers. Table 4.2 presents results as they are reported in the original articles, whereas WHO (1998, Table 4.4) presents relative risks and confidence intervals that have been obtained by further calculation. The calculations used to produce the values in WHO (1998, Table 4.4) are as follows. The risk estimate for a given category of exposure is given by the number of individuals experiencing a particular health outcome divided by the number of individuals in that category. The relative risk for two different exposure categories is given by the ratio of the risks in each category. The variance of the (natural) logarithm of the relative risk is given by the reciprocal of the number of individuals experiencing adverse health effects in the first category, plus the reciprocal of the number of individuals experiencing adverse health effects in the second category, minus the reciprocal of the number of individuals in the first category, minus the reciprocal of the number of individuals in the second category. The standard error of the logarithm of the relative risk is given by the square root of the variance. A 95% confidence interval for the logarithm of the relative risk is obtained by subtracting and adding 1.96 times the standard error to the

logarithm of the relative risk. The anti-logarithm of this 95% confidence interval gives a 95% confidence interval for the relative risk itself. These calculations are appropriate for non-contagious outcomes when the number of individuals in each category is large (Rothman & Greenland, 1998). They are not appropriate when there are confounding factors. In these circumstances, both the estimates of relative risk and their confidence intervals may be misleading.

WHO (1998) is selective as to which results are actually presented: generally only those results which show a positive association between bathing water and health outcome are reproduced. Table 4.2 shows that there is not always a consistent increase in incidence of gastroenteritis with decreasing water quality (as measured by increasing counts of indicator organisms) as would be the case with a well-defined dose-response relationship. There is, however, a general increase in symptom reporting amongst bathers as compared with non-bathers.

Kay *et al.* (1994) report a strikingly higher incidence rate of gastrointestinal symptoms than do other studies for both bathers and non-bathers. Only the results from Fattal *et al.* (1987) for 0-4 year olds (who normally demonstrate a high baseline incidence of gastrointestinal symptoms) approach those obtained by Kay *et al.* (1994). The only other study to present a dose-response relationship between water quality and gastrointestinal symptoms (Cabelli *et al.*, 1982) has been heavily criticised (Fleisher, 1991). Other studies such as Bandaranayake *et al.* (1995) and Haile *et al.* (1999) do not show a consistent increase in gastrointestinal symptoms with decreasing water quality, although the exposure categories in Bandaranayake *et al.* (1995) cover a very limited range of water quality.

4.2 DEGREE OF BIAS

WHO (1998) claims that the UK randomised controlled trials should have the least non-differential misclassification bias. Non-differential misclassification bias applied to measurements of water quality and exposure should, indeed, be smaller in the

randomised controlled trials than in the cohort studies. This is because the UK randomised controlled trials use measurements of water quality at the time and place of exposure, rather than daily or seasonal averages. This does not necessarily explain why the trials identified higher risk estimates for gastroenteric symptoms, nor why Kay *et al.*'s (1994) dose-response curve for gastroenteritis was steeper than those from the other non-UK studies. Higher rates of gastroenteritis in bathers throughout the study might be explained by bias induced by: self-reporting of symptoms; the slightly less stringent definition of gastroenteritis compared with some other studies; the three-week observation period for gastrointestinal symptoms; and recruitment methods (see Section 3.6). The UK randomised controlled trials suffer extensively from both non-differential and differential sources of bias, and this throws into doubt the assertion made by WHO (1998) that non-differential bias is the most important source of bias in the studies they reviewed.

WHO (1998) mentions 'validation of symptoms by medical examination' as a means of reducing bias due to self-reporting of illness. Kay *et al.* (1994) and Fleisher *et al.* (1996) are both cited by WHO (1998) in support of this statement. The wording of WHO (1998) might be taken to imply that self-reported symptoms were verified by medical examination in the studies described in these two articles when in fact they were not. Indeed, the results of the medical examinations conducted as part of UK randomised controlled trials were not in agreement with the self-reported symptoms (see Section 3.7).

WHO (1998) claims that exposure, water quality and illness were much more accurately assessed in the UK randomised controlled trials than in the other studies. It is fair to say that exposure and water quality (on the exposure days) were much more accurately assessed but illness assessment appears to have been seriously biased. WHO (1998) also claims that 'Most causes of bias affecting the selecting studies are likely to lead to underestimation of the health effect of water quality'. In fact, the many sources of bias affecting the design, conduct and analysis of the UK randomised

controlled trials are expected to have led to serious over-estimation of the health effects of faecal pollution (see Section 3.6).

4.3 COMPARISON OF DOSE-RESPONSE CURVES

WHO (1998, Figure 4.3) displays fitted models from various studies of the health effects of sea bathing. The figure is an updated version of a figure in Pike (1991a) and not Pike (1991b) as WHO (1998) claims. The original figure in Pike (1991a) also gives the model formulae used to produce the figure. WHO (1998) omits this useful information and does not explain whether the curves for the models of Kay *et al.* (1994) and Fleisher *et al.* (1996) were plotted using the odds of illness scale (right-hand y-axis) or case rates (left-hand y-axis). The two measures are similar when the risk is small but are not identical and so it would be useful to know which one they have used (or whether they used a mixture).

WHO (1998, Figure 4.3) presents dose-response models for different health outcomes in relation to different bacterial indicators of water quality on the same graph. It is not clear whether direct comparison of the different curves is at all meaningful. It would have been preferable to have presented models for groups of similar health outcomes and comparable indicators on separate graphs.

Kay *et al.*'s (1994) and Fleisher *et al.*'s (1996) dose-response curves are steeper than the curves for the other studies compared in WHO (1998, Figure 4.3). The steepness of the curves from the UK randomised controlled trials might have been caused by lack of control for confounders. It is unlikely to be due to the use of the square root transformation of faecal streptococci counts as all fitted models will pass through the observed data to a greater or lesser degree. It is unlikely to be caused by bias since bathers were blind to water quality (except for knowing that the beaches had met current EC mandatory standards in the previous year).

The overall position of Kay *et al.*'s (1994) and Fleisher *et al.*'s (1996) dose-response curves in WHO (1998, Figure 4.3) could be explained by over-estimation of the risk resulting from sources of bias linked to: self-reporting of symptoms; the definition of gastroenteritis; the observation period for symptoms; and recruitment methods.

4.4 THRESHOLDING

The curvature at the very beginning of the dose-response models of Kay *et al.* (1994) and Fleisher *et al*. (1996) in WHO (1998, Figure 4.3) is due to their imposing a threshold during model fitting. Had no thresholds been imposed, then these logistic models would have implied a straight-line relationship between log-odds of illness and bacterial indicator counts.

WHO (1998) uses the term 'suggested threshold values' to refer to the lowest levels of bacterial counts at which statistically significant elevations in risk of adverse health effects (relative to waters with the lowest bacterial counts) were reported in the articles it reviewed. As mentioned earlier, existence of statistically significant increases in risk does not necessarily imply the existence of clinically important increases, nor a requirement for intervention. These issues can be addressed to a certain extent through risk assessment (see Chapter 5).

The method used by Kay *et al.* (1994) to estimate the threshold in their dose-response model is rather subjective. The estimate of the level of water quality that corresponds to a statistically significant increase in risk (32 faecal streptococci per 100 ml) is, however, consistent with their earlier analyses in which faecal streptococci concentration was treated as a categorical variable (see Section 3.7). The categorical analysis applied to bathers only and has a degree of reliability, as bathers were blind to the exact quality of the water in which they were bathing. The categorical analysis also allowed for a limited set of confounders, whereas the later continuous dose-response model did not allow for any confounders at all. WHO (1998) makes a fair point when it says that the restriction of Kay *et al.*'s (1994) study to healthy adults

could have led to a higher threshold than in other studies. However, the inadequacies in Kay *et al.*'s (1994) model must not be overlooked. In particular, the degree of bias that could have led to over-estimation of risks and the very limited adjustment for confounders make even the analysis treating faecal streptococci as a categorical variable and the resulting threshold of statistically significant increased risk at 40-59 faecal streptococci per 100 ml somewhat unreliable.

4.5 OTHER ISSUES

The issue of sample size is probably not of major importance in the interpretation of the UK randomised controlled trials when they are considered in isolation: if the sample size is too small it can make a true difference impossible to detect but if a statistically significant difference is detected then increasing the sample size is unlikely to make it go away. However, the (unnecessarily) small sample size in the UK randomised controlled trials due to excluding non-bathers hinders comparison with other studies because standard errors and confidence intervals for Kay *et al.*'s (1994) dose-response model could be large relative to other models. Unfortunately, Kay *et al.* (1994) give no indication of the uncertainty attached to their fitted model.

Exclusion of non-bathers from Kay *et al.*'s (1994) final dose-response model also prevents the effects of contact with sea-water itself being investigated. van Dijk *et al.* (1996) found that water-related activity (regardless of water quality) had statistically significant effects on the rates of gastrointestinal symptoms (and other health outcomes) in the UK cohort studies. This type of analysis could have been conducted by Kay *et al.* (1994). Another limitation of Kay *et al.*'s (1994) analysis was the inclusion of subjects who reported additional bathing activity in the period from three days before exposure to three weeks afterwards. Some other studies (for example, Cabelli *et al.*, 1982) excluded such subjects from analysis. These differences between studies and their potential effects on risk estimates are not discussed by WHO (1998).

4.6 SUMMARY

Several studies reviewed by WHO (1998) relate to the UK randomised controlled trials and cohort studies. In particular, the dose-response models of Kay *et al.* (1994) and Fleisher *et al.* (1996) are based on data from the same series of trials and application of very similar methods of statistical analysis. Comparison between the studies reviewed by WHO (1998) should, therefore, take into account the fact that all the dose-response models relating to the UK randomised controlled trials are subject to the sources of bias and inadequacies of statistical analysis highlighted in relation to Kay *et al.*'s (1994) model in Section 3.

The review conducted by WHO (1998) is selective in that results from studies that have not shown associations between adverse health effects and bathing water quality have not been reproduced. The WHO (1998) review is also misleading in that it implies that self-reported symptoms in the UK randomised controlled trials were confirmed by medical examination.

Another drawback of the WHO (1998) review is that estimates of relative risks and confidence intervals have been re-calculated as part of the review, rather than reproduced as they appeared in the original articles. In particular, WHO (1998) presents unadjusted relative risks and confidence intervals (that is, the effects of potential confounders are ignored). The relative risks and confidence intervals presented by WHO (1998) are, therefore, of limited value.

Across all studies, there is a general increase in reporting of gastrointestinal symptoms amongst bathers as compared with non-bathers. However, there is not always a consistent increase in symptom reporting with decreasing water quality as would be expected with a dose-response relationship. Only one non-UK study (Cabelli *et al.*, 1982) presents a dose-response relationship between gastroenteritis and bathing water quality.

Compared with other studies, Kay *et al.* (1994) report a strikingly high rate of incidence of gastroenteritis in both bathers and non-bathers. This could be due to the very substantial sources of bias in the design, conduct and analysis of the UK randomised controlled trials. Also, Kay *et al.*'s (1994) dose-response curve is steeper than the curves from other non-UK studies. This could be due to the absence of control for confounders.

Of all the studies reviewed by WHO (1998), the UK randomised controlled trials probably provide the most precise estimates of water quality and exposure in that measurements of water quality were collected at the time and place of exposure and because exposure on the trial day was closely controlled. The UK randomised controlled trials could also provide acceptable estimates of adverse health effects if the concerns identified in Section 3 were addressed. Further aspects of analysis that were addressed in other studies and which could be applied to advantage in re-analysis of the UK randomised controlled trials are the effects of contact with sea water itself (regardless of water quality) and the exclusion of subjects who reported additional bathing activity. Given re-analysis, the UK randomised controlled trials could provide the most precise estimates of the health effects of sea bathing to date but the present analysis is unacceptable.

Table 4.1 Epidemiological studies of the effects of sea bathing water

Reference	Study dates	Location	Study design	Study size (% after exclusion)	Health outcomes	Indicator organisms (method of analysis)	Comments
Studies using seasonal averages only							
Cabelli (1983)	1976-78	Egypt	Prospective	**1976** 2805 (81.2-88.6) **1977** 7964 (84.8-91.2) **1978** 12311 (84.4-90.6)	Fever Diarrhoea or vomiting Ear/eye/skin Hepatitis Typhoid	Ent *E. coli* (MF)	Head immersion. The study assessed bathing days per week. Differences from US studies (Cabelli *et al.*, 1982) not specified.
Mujeriego *et al.* (1982)	1979	Spain	Retrospective	20918 (not reported)	Skin mycosis Ear infection Throat infection Eye infection Diarrhoea	TC FC FS (NR)	Head immersion.
Studies collecting data daily but analysing averages over study period							
Fattal *et al.* (1987)	May - Aug 1983	Israel	Prospective	2231 (not reported)	Enteric Respiratory Skin disease Ear infection	*E. coli* FC Ent (MF)	Head immersion, swallowing water or face splashed by waves.
Cheung *et al*. (1990)	1987	Hong Kong	Prospective	18741 (77.0)	GI HCGI Ear Eye Skin Respiratory Others	FC *E. coli* *Klebsiella* spp FS Ent Staphylococci *P. aeruginosa* Total fungi (MF)	Head immersion or face splashed.

Reference	Study dates	Location	Study design	Study size (% after exclusion)	Health outcomes	Indicator organisms (method of analysis)	Comments
Pike (1994) van Dijk *et al.* (1996)	1989-92	UK	Prospective	16569 (78 overall; 92 for 9 beaches and 20 for 1 beach; figures extracted from Pike, 1994)	GI Diarrhoea Respiratory Ear, nose and throat Eye Skin	TC FC FS Coliphages[a] (MF)	Sampled water at 'mid' depth every 2 hours on day of interview but water quality was not related to individuals' exposure. Those entering the water could have multiple exposures. Specified wading, bathing and surfing/diving but did not define bathing.
Studies using daily measurements							
Cabelli *et al.* (1982)	1973-78	USA	Prospective	**1973** 4754 (77.2-86.6) **1974** 11388 (77.9-82.9) **1975** 6491 (78.3) **1978** 4053 (81.2)	GI (HCGI) Respiratory Other disabling	*E. coli* FC Ent *Klebsiella* spp TC (MF)	Head immersion (determined by inquiry and observation).
Mariño *et al.* (1995)	1988-90	Spain	Prospective	9691 (not reported)	Enteric (GI) Otitis Conjunctivitis Upper respiratory Dermatitis Urinogenital	TC FC *E. coli* FS *P. aeruginosa* Coliphages (NR)	Head immersion.
Corbett *et al.* (1993)	1989-90	Australia	Prospective	2968 (129 subjects excluded because of missing details)	GI Respiratory Fever Eye Ear Any	FC FS (NR)	Face and head immersion. Duration in water estimated.

Reference	Study dates	Location	Study design	Study size (% after exclusion)	Health outcomes	Indicator organisms (method of analysis)	Comments
Pike (1994) Kay *et al*. (1994) Fleisher *et al.* (1996)	1989-92	UK	Randomised controlled trials	1216 (93.1; 104 subjects excluded because of GI on day of trial; 1112 analysed)	Objective GI Subjective GI Respiratory Ear/eye Skin	TC FC FS Staphylococci *P. aeruginosa* (MF)	Head immersion (3 times) and at least 10 minutes in the water. Water samples taken every 20-30 minutes and personal exposure determined.
Kueh *et al*. (1995)	1992	Hong Kong	Prospective	18122 (72.4)	GI HCGI Respiratory HC respiratory Eye Skin	*E. coli* FC Staphylococci (NR)	Bathing specified as getting the face wet.
Bandaranayake *et al.* (1995)	1995	New Zealand	Prospective	3887 (70.3)	GI HCGI Respiratory Ear/eye/throat/skin	FC *E. coli* Ent (MF)	Head immersion. Assessment of time spent bathing.
Wang (1998) Haile *et al.* (1999)	1995	USA	Prospective	10459 (78.8; figures extracted from Haile *et al.*, 1999)	HCGI HCGI 2[b] Significant respiratory disease Eye/ear/skin Others	TC FC *E. coli* Ent (MF)	Head immersion.
Nelson & Williams (1997)	1995-96	Wales, UK	Prospective	580 (phone nos) 535 (contacted)	GI Ear problems Any illness	FC FS (MF)	'Differing levels of immersion'. Assessed confounders. Bathers at higher risk but no dose-response relationship. Low numbers in survey. Not clear how health effects were related to water quality.

Ent, enterococci; FC, faecal coliforms; FS, faecal streptococci; GI, gastrointestinal illness; HCGI, highly credible GI; MF, membrane filtration; NR, microbiological method of analysis not reported; TC, total coliforms

[a] Microbiological method of analysis for coliphages not reported; [b] Vomiting and fever

Table 4.2 Studies relating gastrointestinal symptoms to faecal streptococci, enterococci, *E. coli* and faecal coliforms

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Studies using faecal streptococci and enterococci						
Cabelli *et al.* (1982)	Ent	GI	**New York**			No account taken of confounders.
			1973			
			1.2-59 non-bathers		46	Only GI deemed to be bathing and pollution related.
			1.2-59 bathers		81	
			6-186 non-bathers		24	
			6-186 bathers		72	Different trial days clustered to give exposure category range.
			1974			
			2-5 non-bathers		23	
			2-5 bathers		27	
			7 non-bathers		34	
			7 bathers		38	
			10-17 non-bathers		17	
			10-17 bathers		42	
			30-33 non-bathers		23	
			30-33 bathers		43	
			1975			
			2-11 non-bathers		55	
			2-11 bathers		63	
			14-38 non-bathers		37	
			14-38 bathers		59	
			86-298 non-bathers		31	
			86-298 bathers		60	
			Boston			
			1978			
			2-6 non-bathers		66	
			2-6 bathers		83	
			6-9 non-bathers		67	
			6-9 bathers		71	
			12 non-bathers		74	
			12 bathers		108	

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Cabelli *et al.* (1982)		HCGI	**New York**			
			1973			
			1.2-59 non-bathers		15.2	
			1.2-59 bathers		30.4	
			6-186 non-bathers		18.0	
			6-186 bathers		46.4	
			1974			
			2-5 non-bathers		4.2	
			2-5 bathers		7.6	
			7 non-bathers		6.9	
			7 bathers		10.5	
			10-17 non-bathers		2.4	
			10-17 bathers		16.0	
			30-33 non-bathers		Not reported	
			30-33 bathers		18.1	
			1975			
			2-11 non-bathers		19.3	
			2-11 bathers		18.8	
			14-38 non-bathers		7.4	
			14-38 bathers		14.8	
			86-298 non-bathers		Not reported	
			86-298 bathers		34.5	
			Boston			
			1978			
			2-6 non-bathers		11.0	
			2-6 bathers		23.0	
			6-9 non-bathers		28.0	
			6-9 bathers		33.0	
			12 non-bathers		13.0	
			12 bathers		41.0	

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Cabelli (1983)	Ent	Diarrhoea/ vomiting	**Alexandria residents**			Unclear as to whether or not confounders were addressed. Differences from US studies (Cabelli *et al.*, 1982) not specified.
			Maamoura			
			1976			
			103 non-bathers		11.5	
			103 bathers		16.1	
			1977			
			72.8 non-bathers		8.5	
			72.8 bathers		12.2	
			1978			
			214 non-bathers		6.5	
			214 bathers		10.3	
			Ibraheima			
			1976			
			286 non-bathers		3.2	
			286 bathers		15.6	
			1977			
			211 non-bathers		2.0	
			211 bathers		16.2	
			1978			
			954 non-bathers		13.0	
			954 bathers		21.1	
			Mandara			
			1976			
			5760 non-bathers		17.6	
			5760 bathers		31.3	
			Sporting			
			1977			
			6780 non-bathers		5.2	
			6780 bathers		29.6	
			1978			
			9160 non-bathers		7.8	
			9160 bathers		19.2	

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Cabelli (1983)	Ent	Diarrhoea/ vomiting	**Cairo visitors**			.
			Maamoura			
			1977			
			72.8 non-bathers		22.6	
			72.8 bathers		21.5	
			1978			
			214 non-bathers		18.1	
			214 bathers		17.5	
			Ibraheima			
			1977			
			211 non-bathers		13.0	
			211 bathers		25.9	
			1978			
			954 non-bathers		7.5	
			954 bathers		48.3	
			Sporting			
			1977			
			6780 non-bathers		12.4	
			6780 bathers		51.2	
			1978			
			9160 non-bathers		7.2	
			9160 bathers		44.8	
Fattal *et al*. (1987)	Ent	Enteric disease in 0-4 year olds	0-24 non-bathers		90	No account taken of confounders.
			0-24 bathers		114	
			25-410 non-bathers		133	
			25-410 bathers		221	

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Cheung *et al.* (1990)	Ent	GI	Non-bathers		Not reported	'Incidence rates' are swimming-associated symptoms rates (incidence rates for bathers minus incidence rates for non-bathers).
			0-39		3.6	
			40-250		4.5	
		HCGI	Non-bathers		Not reported	
			0-39		2.2	Details of pre-trial illness, bathing activity and food eaten not incorporated into any calculations/conclusions.
			40-250		1.9	
	FS	GI	Non-bathers		Not reported	
			0-55		1.8	
			56-290		5.1	
		HCGI	Non-bathers		Not reported	
			0-55		0.9	
			56-290		2.5	
Kay *et al.* (1994)	FS	Objective gastroenteritis	Non-bathers		97	Relative risk is odds ratio relative to 0-39 exposure category. Account taken of non-water related risk factors.
			0-19		109	
			20-39		106	
			40-59	1.91	183	
			60-79	2.90	281	
			80+	3.17	304	
Bandaranayake *et al.* (1995)	Ent	GI	Non-bathers		35	Relative risks adjusted for age group.
			0-1.5	1.18	55	
			1.5-3.75	1.13	55	
			3.75-13	1.43	64	
			13-232.25	1.44	66	
		HCGI	Non-bathers		17	
			0-1.5	1.13	26	
			1.5-3.75	0.84	18	
			3.75-13	1.00	16	
			13-232.25	1.38	22	
Haile *et al.* (1999)	Ent	HCGI 1[b]	0-35		30	No non-bathers.
			35-104	0.92 (0.67-1.26)	27	Relative risk in relation to lowest exposure group (adjusted for age, sex, race, resident-or-not and concern about beach hazards).
			104+	1.31 (0.89-1.92)	42	
		HCGI 2[c]	0-35		9	
			35-104	0.82 (0.46-1.48)	8	
			104+	1.30 (0.67-2.51)	14	

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Studies using *E. coli* and faecal coliforms						
Cabelli *et al.* (1982)	*E. coli*	GI	**New York**			No account taken of confounders (assumed to be included in non-bather rates).
			1973			
			3-34 non-bathers		46	Results from part of the study are not reproduced here as they relate to non-marine waters.
			3-34 bathers		81	
			50-708 non-bathers		24	
			50-708 bathers		72	
			1974			Only GI deemed to be bathing-and-pollution-related.
			1-4 non-bathers		34	
			1-4 bathers		25	
			9-19 non-bathers		29	
			9-19 bathers		38	
			26-35 non-bathers		53	
			26-35 bathers		65	
			1975			
			22-89 non-bathers		51	
			22-89 bathers		55	
			115-169 non-bathers		41	
			115-169 bathers		76	
			208-356 non-bathers		24	
			208-356 bathers		55	
			441-659 non-bathers		55	
			441-659 bathers		68	
			Boston			
			1978			
			4-7 non-bathers		63	
			4-7 bathers		72	
			5-9 non-bathers		68	
			5-9 bathers		86	
			13-22 non-bathers		67	
			13-22 bathers		70	
			28-31 non-bathers		71	
			28-31 bathers		93	

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Cabelli *et al.* (1982)	*E. coli*	HCGI	**New York**			
			1973			
			3-34 non-bathers		15.2	
			3-34 bathers		30.4	
			50-708 non-bathers		18.0	
			50-708 bathers		46.4	
			1974			
			1-4 non-bathers		3.7	
			1-4 bathers		8.0	
			9-19 non-bathers		5.7	
			9-19 bathers		14.1	
			26-35 non-bathers		2.4	
			26-35 bathers		23.3	
			1975			
			22-89 non-bathers		17.8	
			22-89 bathers		13.4	
			115-169 non-bathers		10.3	
			115-169 bathers		24.5	
			208-356 non-bathers		3.0	
			208-356 bathers		21.0	
			441-659 non-bathers		7.4	
			441-659 bathers		24.5	
			Boston			
			1978			
			4-7 non-bathers		29.0	
			4-7 bathers		39.0	
			5-9 non-bathers		10.0	
			5-9 bathers		23.0	
			13-22 non-bathers		27.0	
			13-22 bathers		27.0	
			28-31 non-bathers		14.0	
			28-31 bathers		32.0	
Cabelli *et al.* (1982)	FC					Poor correlation with bathing-associated GI symptoms

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Cabelli (1983)	*E. coli*	Diarrhoea/ vomiting	**Alexandria residents**			
			Maamoura			
			1976			
			14.6 non-bathers		11.5	
			14.6 bathers		16.1	
			1977			
			35.3 non-bathers		8.5	
			35.3 bathers		12.2	
			1978			
			53.1 non-bathers		6.5	
			53.1 bathers		10.3	
			Ibraheima			
			1976			
			184 non-bathers		3.2	
			184 bathers		15.6	
			1977			
			415 non-bathers		2.0	
			415 bathers		16.2	
			1978			
			668 non-bathers		13.0	
			668 bathers		21.1	
			Mandara			
			1976			
			1620 non-bathers		17.6	
			1620 bathers		31.3	
			Sporting			
			1977			
			6300 non-bathers		5.2	
			6300 bathers		29.6	
			1978			
			10400 non-bathers		7.8	
			10400 bathers		19.2	

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Cabelli (1983)	*E. coli*	Diarrhoea/ vomiting	**Cairo visitors**			
			Maamoura			
			1977		22.6	
			35.3 non-bathers		21.5	
			35.3 bathers			
			1978		18.1	
			53.1 non-bathers		17.5	
			53.1 bathers			
			Ibraheima			
			1977		13.0	
			415 non-bathers		25.9	
			415 bathers			
			1978		7.5	
			668 non-bathers		48.3	
			668 bathers			
			Sporting			
			1977		12.4	
			6300 non-bathers		51.2	
			6300 bathers			
			1978			
			10400 non-bathers		7.2	
			10400 bathers		44.8	
Fattal *et al.* (1987)	FC	Enteric diseases in 0-4 yr olds	0-50 non-bathers			No account taken of confounders.
			0-50 bathers			
			51-650 non-bathers			Incidence rates reported graphically.
			51-650 bathers			
	E. coli	Enteric diseases in 4 yr olds	0-24 non-bathers			No excess relative risk in bathers.
			0-24 bathers			
			25-268 non-bathers			Incidence rates reported graphically.
			25-268 bathers			

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Cheung *et al.* (1990)	FC	GI	Non-bathers		Not reported	'Incidence rates' are swimming-associated symptom rates (incidence rates for bathers minus incidence rates for non-bathers).
			0-410		1.8	
			411-3200		5.1	
		HCGI	Non-bathers		Not reported	
			0-410		0.9	Details of pre-trial illness, bathing activity and food eaten not incorporated into any calculations/conclusions.
			411-3200		2.5	
	E. coli	GI	Non-bathers		Not reported	
			0-180		2.8	
			181-1800		5.3	
		HCGI	Non bathers		Not reported	
			0-180		1.1	
			181-1800		2.6	
Kueh *et al.* (1995)	*E. coli*	GI	Bathers (B1)		2.0	Symptom rates (corrected for non-bather rates) given for two beaches having seasonal geometric means of 13 *E. coli* per 100 ml (B1); and 237 *E. coli* per 100 ml (B2). No account taken of confounders.
			Bathers (B2)		9.0	
		HCGI	Bathers (B1)		0.1	
			Bathers (B2)		6.2	
	FC		Not reported		Not reported	No direct relationship between GI and *E. coli* or faecal coliforms could be identified in this study.

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Bandaranayake *et al.* (1995)	FC	GI	0.375-3.5	1.53		Risk ratios adjusted for age group.
			3.5-8	0.99		
			8-35	1.69		
			35-724	0.83		
		HCGI	0.375-3.5	1.37		
			3.5-8	0.89		
			8-35	1.05		
			35-724	0.93		
	E. coli	GI	0-3	1.63		
			3-6	0.78		
			6-28.25	1.74		
			28.25-552	0.94		
		HCGI	0-3	1.48		
			3-6	0.72		
			6-28.25	1.16		
			28.25-552	0.80		
Haile *et al.* (1999)	FC	HCGI	0-200		30	No non-bathers. Relative risk was calculated in relation to lowest exposure group (adjusted for age, sex, race, resident-or-not & concern about beach hazards).
			200-400	1.18 (0.79-1.77)	36	
			400+	0.99 (0.72-1.36)	31	
		HCGI 2[b]	0-200		8	
			200-400	1.63 (0.85-3.12)	14	
			400+	1.13 (0.65-1.95)	10	
	E. coli	HCGI	0-35		30	
			35-75	1.03 (0.75-1.42)	31	
			75-160	0.88 (0.59-1.30)	26	
			160-320	1.06 (0.63-1.80)	33	
			320+	1.12 (0.76-1.64)	36	
		HCGI 2[b]	0-35		8	
			35-75	1.55 (0.92-2.64)	13	
			75-160	0.85 (0.40-1.81)	7	
			160-320	1.25 (0.51-3.03)	12	
			320+	1.04 (0.51-2.13)	10	

Ent, enterococci; FC, faecal coliforms; FS, faecal streptococci; GI, gastrointestinal illness; HCGI, highly credible GI

[a] Exposure categories for non-bathers refer to levels of exposure by bathers on the same beach as the non-bathers (organisms per 100ml); [b] Vomiting and fever

5 Use of the results of the UK randomised controlled trials in risk assessment

5.1 REQUIREMENTS OF AN IDEAL RISK ASSESSMENT STRATEGY

The following issues require consideration in the development of an ideal approach to risk assessment relating health effects to environmental exposure.

- Characterisation of exposure.
- Characterisation of health effects.
- Existence of a dose-response relationship quantifying the relationship between the level of exposure and health effects.
- Characterisation of the degree of uncertainty in risk estimates. Uncertainty relates to the estimation of risks from a finite sample, rather than the entire population of interest, and is characterised by standard errors or confidence intervals.
- Characterisation of all sources of variation in risk (these include the exposure measurement itself, potential confounders, such as age and sex, and other variables, such as geographical location), and presentation of the results of risk assessment at a disaggregated level (for example, separate estimates for different age groups and sexes).
- Sensitivity analysis (to examine the effects of model assumptions on estimates of risk).

These issues are highlighted in various publications relating to environmental risk assessment (for example, Covello & Merkhofer, 1993; Barnett & O'Hagan, 1997; Nurminen *et al*., 1999), although some issues are of more general concern. Other issues that may require consideration are the determination of tolerable levels of risk and the examination of costs of remedial action needed to control exposure levels. However,

these are largely policy issues, and as such relate more to risk management than to risk assessment.

5.2 RISK ASSESSMENT STRATEGY USED IN THE WHO GUIDELINE DOCUMENT

Overview

The main elements of the risk assessment strategy used by WHO (1998) are as follows.

- Characterisation of exposure using faecal streptococci as an indicator of water quality.
- Characterisation of health effects in terms of gastrointestinal symptoms.
- Adoption of Kay *et al.*'s (1994) dose-response model quantifying the relationship between gastroenteritis and water quality.
- Assumption that $\log_{10}$ faecal streptococci counts are normally distributed with a standard deviation of 0.8103.
- Criterion for compliance with microbiological standards set to be that 95% of water samples will result in faecal streptococci counts equal to or less than the standard.
- Calculation of the mean $\log_{10}$ faecal streptococci count (and hence the 95 percentile to be proposed as a standard, or guideline value) required for the expected proportion of exposures that lead to gastroenteritis to be less than a chosen target value.
- Comparison of guideline values for faecal streptococci counts corresponding to different choices for the target value.

The reasons for choosing faecal streptococci as an indicator of faecal pollution, gastroenteritis as an appropriate adverse health effect, and Kay *et al.*'s (1994) dose-

response model were discussed in Sections 2, 3 and 4. The details of the calculation of the guideline values are discussed in the remainder of Section 5.

Calculation of guideline values

The WHO's (1998) calculation of guideline values is based on the expected proportion of exposures that lead to gastrointestinal symptoms. The expected proportion of exposures that lead to gastrointestinal symptoms appears to be have been obtained by:

- drawing 1000 random variables from the distribution of $\log_{10}$ faecal streptococci counts;
- calculating the probability that exposure leads to gastrointestinal symptoms for each faecal streptococci count using Kay *et al.*'s (1994) dose-response model expressed in terms of excess risk due to bathing;
- summing the probabilities so obtained to yield the expected number of exposures that lead to gastrointestinal symptoms; and
- dividing by the number of exposures (1000).

In order to perform the above calculations, the mean and standard deviation of the distribution of $\log_{10}$ faecal streptococci counts must be specified. The standard deviation is fixed at 0.8103. This value is regarded as typical by the WHO (1998), being derived from a study of more than 11000 EC bathing waters (see CREH, 1996). The mean of the distribution is chosen to be the largest value for which the expected proportion of exposures that lead to gastrointestinal symptoms is less than a certain target value. Once the mean $\log_{10}$ faecal streptococci count has been determined, the 95 percentile of the distribution (the $\log_{10}$ faecal streptococci count below which 95% of observations will fall) can be calculated. The guideline value is the 95 percentile expressed as a count (rather than a $\log_{10}$ count) per 100 ml.

The calculation of the expected proportion of exposures that lead to gastrointestinal symptoms could have been performed analytically, that is, using mathematical calculus, in which case the act of summing over the probabilities would be described as integrating the probabilities over the distribution of $\log_{10}$ faecal streptococci counts. The method used by WHO (1998) appears, however, to be based on a form of numerical integration called Monte Carlo simulation (see, for example, Covello & Merkhofer, 1993; Barnett & O'Hagan, 1997; Nurminen *et al.*, 1999). The expected risk of adverse health outcomes calculated by integrating a dose-response curve over the distribution of exposure measurements (sometimes for different subgroups of the population—see below) is called the disease burden (see, for example, Nurminen *et al.*, 1999).

The first target value used by the WHO (1998) is 0.05 (corresponding to a single excess incidence of gastrointestinal symptoms in 20 exposures). The reason for focusing on 20 exposures is that the WHO (1998) assumes that a 'typical bather' receives 20 exposures during a bathing season (for example, by bathing twice a day during a ten-day holiday or by bathing twice a day on 10 visits to the beach spread out over the bathing season). Since the WHO (1998) risk assessment model is based on Kay *et al.*'s (1994) dose-response model, the typical bather corresponds to the average bather in the UK randomised controlled trials for whom a single exposure corresponds to at least 10 minutes of bathing in which head immersion occurs at least three times. The $\log_{10}$ faecal streptococci counts experienced by the same bather on separate exposure occasions are assumed to be statistically independent. The WHO (1998) guideline value based on an individual bather receiving 20 exposures over a single bathing season is 200 faecal streptococci per 100 ml.

WHO (1998) also considers risks to groups of individuals experiencing gastrointestinal symptoms as a result of sea bathing. The group considered is a family of four healthy adult bathers. Each individual is assumed to receive 20 exposures per bathing season. The $\log_{10}$ faecal streptococci counts to which different individuals are exposed are

assumed to be statistically independent. The target value for this scenario is 0.0125 (corresponding to a single excess incidence of gastrointestinal symptoms in 80 exposures). The WHO (1998) guideline value based on a family of four healthy adults each receiving 20 exposures over a single bathing season is 50 faecal streptococci per 100 ml.

WHO (1998) further considers risks to a family of four healthy adult bathers over five bathing seasons. The $\log_{10}$ faecal streptococci counts corresponding to different seasons are assumed to be statistically independent. The target value for this scenario is 0.0025 (corresponding to a single excess incidence of gastrointestinal symptoms in 400 exposures). The WHO (1998) guideline value based on a family of four healthy adults, each receiving 20 exposures per bathing season over five bathing seasons is 10 faecal streptococci per 100 ml.

WHO (1998, Table 4.7) compares the different guideline values with thresholds of risk reported in the articles they reviewed. The guideline value of 10 faecal streptococci per 100 ml is below the 'no observed adverse effect level' in most epidemiological studies that have attempted to define one. The guideline value of 50 faecal streptococci per 100 ml is above the 'lowest observed adverse effect level' for gastroenteritis in most epidemiological studies that have attempted to define one. The guideline value of 200 faecal streptococci per 100 ml is above the lowest observed adverse effect level for all adverse health outcomes in most epidemiological studies.

A fourth guideline value proposed by WHO (1998) is 1000 faecal streptococci per 100 ml. This value corresponds to a concentration of total coliforms thought to be associated with risks of transmission of severe health outcomes such as typhoid fever (PHLS, 1959). WHO (1998) derived this value by converting total coliform counts to faecal coliform counts, and then converting faecal coliform counts to faecal streptococci counts.

5.3 LIMITATIONS OF THE WHO RISK ASSESSMENT STRATEGY

Overview

WHO's (1998) overall approach to risk assessment appears to be reasonable, and its characterisation of exposure and health effects seems to be appropriate. There are however, many limitations in its approach, including:

- reliance on Kay *et al.*'s (1994) dose-response model, which is subject to bias, inadequate control for confounders and other limitations;
- failure to characterise the degree of uncertainty in risk estimates;
- failure to characterise all sources of variation in risk and to present the results of risk assessment in disaggregated form; and
- failure to undertake a sensitivity analysis to examine the effects of model assumptions relating to the standard deviation of $\log_{10}$ faecal streptococci counts and independence between exposures on different occasions and amongst family groups.

These and other issues (including the range of faecal streptococci concentrations to which bathers were exposed in the UK randomised controlled trials, and the definition of target values) are discussed in the remainder of this section.

Impact of limitations of the dose-response model

Since Kay *et al.*'s (1994) dose-response model suffers from bias and inadequate control for confounders, these characteristics will be inherited by the guideline values. The proposed guideline values are therefore expected to be over-estimates of the degree of water quality required to meet the target values. Comparison of the proposed guideline values with thresholds of statistically significant elevations of risk reported in other articles does not necessarily confirm the reliability of the guideline values. The risk assessment based on Kay *et al.*'s (1994) dose-response model has the

potential to yield much more meaningful water quality standards than analyses based solely on statistically significant increases in risk. This is because the risk assessment strategy incorporates a measure of disease burden and attempts to define the clinical importance of risk in terms of target values (see below). Improved guideline values could be obtained after re-analysis of the data from the UK randomised controlled trials to yield a dose-response model that controls for all possible confounders. Such a model could be used either to standardise risk estimates to reflect, for example, a particular age distribution (not necessarily the same as that in the study population for the UK randomised controlled trials). Estimates of this type could be used to investigate population attributable risk, that is, the level of risk associated with the current pattern of exposure to sea bathing water in the general population (see, for example, Rothman & Greenland, 1998). Alternatively, a dose-response model that takes full account of all possible confounders could be used to provide separate estimates of water quality standards required to protect, for example, bathers of different ages. Separate estimates would allow variations in risk due to different sources of variation to be compared (see below).

WHO (1998) claims that it is 'appropriate to express the health risk in terms of excess risk of illness to the exposed relative to the unexposed'. From a practical point of view, this approach is sensible. However, Kay *et al.*'s (1994) dose-response model is based on an assumption that risks are multiplicative, rather than additive (see Section 3.7). Potential bias in the baseline risk (that is, the risk to non-bathers) of experiencing gastroenteritis could lead to under- or over-estimation of water quality standards.

An alternative approach would be to estimate risks linked to a particular distribution of faecal streptococci counts for bathers and non-bathers separately (rather than for the difference between bathers and non-bathers). Monte Carlo simulation would lend itself to more complex modelling scenarios, including separate estimation of risks for bathers and non-bathers. The advantage of having these separate estimates is that, as well as eventually allowing the difference in risks to be evaluated, the effects on risk of

bias, uncertainty, and variation due to potential confounders could be examined separately for bathers and non-bathers. The effect of potential bias in the estimates of baseline risk (to non-bathers) in the UK randomised controlled trials could itself be examined using Monte Carlo simulation.

Uncertainty in risk estimates

Two major sources of uncertainty in the WHO (1998) risk assessment model are the precision of the estimated standard deviation of $\log_{10}$ faecal streptococci counts and the precision of Kay *et al.*'s (1994) dose-response model. The precision of Kay *et al.*'s (1994) dose-response model could be low because of the relatively small sample size and the restriction to bathers. It should be noted that uncertainties in different elements of a complex model will be compounded, giving rise to greater uncertainties overall. The combined effects of sources of uncertainty such as these could be examined using Monte Carlo simulation or other appropriate statistical techniques (see, for example Covello & Merkhofer, 1993; Barnett & O'Hagan, 1997; Nurminen *et al.*, 1999). Another source of uncertainty identified by WHO (1998) which should be examined is that resulting from imprecision of current techniques for enumerating indicator organisms. The aim of uncertainty analysis should be to provide estimates of the precision of guideline values, whether in the form of standard errors or confidence intervals. It should be noted, however, that it is not acceptable to use the product of confidence limits for individual components of uncertainty in order to estimate overall confidence limits. Nor can the confidence intervals given in WHO (1998, Table 4.4) be used as surrogates for the true level of uncertainty in Kay *et al.*'s (1994) dose-response model.

Sources of variation

The major sources of variation which should be examined and quantified are those due to potential confounders and to geographical locations. Extension of Kay *et al.*'s (1994) dose-response model to allow for such factors (and to include non-bathers)

would allow these effects to be quantified as part of risk assessment using, for example, Monte Carlo simulation. It would be possible to present guideline values for more representative groups of individuals (relative to the entire population) than those in the UK randomised controlled trials, and these results could be presented in disaggregated form to allow the effects of different sources of variation in risk to be compared.

Sensitivity analysis

The calculation of guideline values is based on assumptions of independence of risks over different exposure occasions for the same individual, and between different individuals in a family group. Sensitivity analysis could be performed to examine the effects of these assumptions. In particular, the assumption of independence between members of the same family could be relaxed using estimated effects of within-family transmission of gastroenteritis in a revision of Kay *et al.*'s (1994) dose-response model. If the uncertainty of the estimated standard deviation of $\log_{10}$ faecal streptococci counts cannot be estimated, the sensitivity of the guideline values to the assumed value for the standard deviation could instead by examined by sensitivity analysis. This could be achieved by substituting standard deviations for individual EC Member States (see CREH, 1996) for the average standard deviation over all Member States that is currently used. Another aspect of the WHO (1998) risk assessment model that could be examined by sensitivity analysis is the impact of potential bias in estimates of the baseline risk (to non-bathers) of experiencing gastrointestinal symptoms. The effects of relying on Kay *et al.*'s (1994) dose-response model could also be examined by substituting a different dose-response model linking gastroenteritis to faecal streptococci or enterococci, for example, Cabelli *et al.'s* (1982) dose-response model. One way of performing the sensitivity analyses outlined above would be to use Monte Carlo simulation.

Other issues

For ethical reasons, the UK randomised controlled trials were restricted to bathing waters that had met current EC mandatory standards in the year preceding each trial. This might have resulted in subjects in the UK randomised controlled trials being exposed to a smaller range of faecal streptococci concentrations than in other studies. Risks for high faecal streptococci counts may, therefore, be poorly estimated in Kay *et al.*'s (1994) dose-response model. It is difficult to quantify the impact of this effect on the guideline values derived by WHO (1998). However, WHO (1998) correctly observes that 'It is wise not to extrapolate beyond the range of values from which the dose-response curve was derived'. It is not clear exactly how the risk of experiencing gastrointestinal symptoms was evaluated in the WHO (1998) risk assessment model for $\log_{10}$ faecal streptococci counts equivalent to 200 faecal streptococci per 100 ml since the highest faecal streptococci count reported in Kay *et al.* (1994, Table 4) is 158.

A further potential limitation of WHO's (1998) risk assessment model is the definition of target values. WHO (1998) argues that 'acceptable' or 'tolerable' excess disease rates 'are especially controversial because of the voluntary nature of recreational water [activity] and the generally self-limiting nature of the most commonly studied health outcomes'. The intention of WHO (1998) is clearly to present guideline values corresponding to various different scenarios of clinical importance. The weight to be given to the different levels of clinical importance (assuming the guidelines are re-calculated according to the suggestions made here) requires consideration. Finally, it should be noted that re-calculation of guideline values using the approaches indicated will not address issues such as variation in risk due to different sources of faecal pollution, degree and type of treatment applied to sewage, effects of rainfall, wind, tides and currents and coastal physiography (see Section 2), nor the differential die-off rates between faecal streptococci and the pathogens they are intended to represent. These limitations are clearly identified by WHO (1998).

5.4 SUMMARY

The approach to risk assessment adopted by WHO (1998) incorporates certain desirable elements, namely: characterisation of exposure (in terms of faecal streptococci counts); characterisation of adverse health effects (in terms of gastrointestinal symptoms); and an attempt to use a dose-response model that links exposure to health effects in order to estimate disease burden. However, it falls some way short of an acceptable risk assessment in that it fails to characterise uncertainty and variability in the proposed guideline values, and because no form of sensitivity analysis is undertaken. The risk assessment could be extended to address these issues using Monte Carlo simulation, or indeed any other appropriate statistical technique (for example, Bayesian inference). Given the reliance on Kay *et al.*'s (1994) dose-response model, the current guideline values are expected to be over-estimates of the degree of water quality required to protect human health. In addition to the extensions to the WHO (1998) risk assessment listed above, re-analysis of the UK randomised control trial data will, therefore, be required in order to derive meaningful guideline values.

6 Recommendations

6.1 OVERVIEW

This review has identified two major areas of concern in the analysis presented in WHO (1998). These are: the statistical analysis of the UK randomised controlled trials to yield Kay *et al.*'s (1994) dose-response model linking gastrointestinal symptoms to faecal streptococci counts; and the WHO risk assessment strategy used to derive guideline microbiological standards. Both areas of concern throw into doubt the validity and reliability of the proposed guideline values, and both areas require re-analysis in order to allow satisfactory estimates of guideline values to be obtained. The specific requirements in each area are summarised below.

6.2 STATISTICAL ANALYSIS OF THE UK TRIAL DATA

The minimum requirements for re-analysis of the UK randomised controlled trial data are as follows. The incidence of gastrointestinal symptoms (in both bathers and non-bathers) in relation to faecal streptococci counts and other factors should be examined using logistic regression models. The response variable (incidence of gastrointestinal symptoms) should be defined to reflect both the severity of symptoms used to define gastroenteritis in other studies (for example, Cabelli *et al.*, 1982) and the expected latency period. Sensible choices for the latency period would be:

- symptoms recorded within seven days of exposure;
- symptoms recorded between eight and 14 days after exposure; and
- symptoms recorded between 15 and 21 days after exposure.

Models for different response variables should be compared with each other and with results from other studies.

The following explanatory variables should be included in the models.

- Exposure status (bather/non-bather) represented as an indicator variable. Inclusion of such a term would allow the effects of sea water alone (irrespective of the degree of pollution) to be investigated. This aspect is lacking from Kay *et al.*'s (1994) dose-response models but could be tackled using van Dijk *et al.*'s (1996) approach, that is by fitting the indicator variable to estimate the average effect of contact with sea water, plus an interaction between the indicator and faecal streptococci counts (see below) to estimate the difference between the average effect and the effect at different levels of water quality.
- Geographical location (trial site) represented as a factor or a random effect.
- Potential confounders:

  - age;
  - sex;
  - frequency of diarrhoea;
  - illness in the four weeks preceding exposure day that lasted for more than 24 hours;
  - consumption of foods associated with gastroenteritis;
  - illnesses in the household between exposure and the end of symptom recording;
  - additional bathing in the period from three days before exposure to the end of symptom recording;
  - awareness of beach maintenance in the UK;
  - awareness of pollution;
  - awareness of news/media coverage of the trials;
  - holidays/business trips in the UK and abroad in the period from four weeks before exposure to the end of symptom recording;
  - duration of exposure; and

- ingestion of water.

Representation of one or more confounders as a single composite variable should be avoided.

- Faecal streptococci counts represented as a continuous variable.

Terms that are not measurements of exposure should be fitted first, so that variability in risks due to these sources can be removed before examining the effects of water quality, etc. The order in which terms are fitted will not affect estimates of regression coefficients but it will affect the analysis of deviance table. The terms in this table should be examined to compare the strength of the effects of different confounding variables and exposure. Standard errors or confidence intervals for risk estimates, odds ratios, or estimated effects of individual explanatory variables should be reported. Goodness-of-fit of the models should also be discussed.

The analysis should include an assessment of the form of the continuous variable representing faecal streptococci counts that gives the best fit. It would also be advisable to explore the possibility of interactions between faecal streptococci counts and factors such as geographical location. The evidence for a non-linear relationship between log-odds of gastrointestinal symptoms and faecal streptococci counts could also be explored by fitting a quadratic term in faecal streptococci. Fitted models obtained using the methods outlined above would yield estimates of risks of experiencing gastroenteritis for both bathers and non-bathers, at a given level of water quality and at given levels of other factors in the models. These estimates, and the corresponding standard errors or confidence intervals, could then be fed into the WHO risk assessment.

Re-analysis according to these specifications will ameliorate bias due to the definition of gastroenteritis, the three-week observation period, repeated statistical testing and model-selection bias. Additional sources of bias that are inherent in the study design but cannot be quantified by re-analysis are:

- restriction to healthy adult volunteers;
- self-reporting of symptoms;
- length of recall period;
- recruitment methods; and
- timing of exposure-day interviews.

The interpretation of the results obtained by re-analysis should attempt to incorporate qualitative statements regarding the expected effects of these sources of bias.

6.3 WHO RISK ASSESSMENT STRATEGY

The minimum requirements for re-analysis of risk are to incorporate uncertainty, variability, and sensitivity analysis as detailed below. All of these requirements could be addressed using Monte Carlo simulation as indicated in Section 5. This is not, however, the only valid method for conducting the analysis required and re-analysis by any appropriate method (for example, Bayesian inference) would be acceptable.

The analysis of uncertainty in risk estimates should address both the uncertainty attached to the re-fitted dose-response model, and uncertainty due to the unknown distribution of $\log_{10}$ faecal streptococci counts.

The major sources of variation in risk that should be examined and quantified are those due to potential confounders and geographical locations. After re-analysis of the UK randomised controlled trial data, it should be possible to present guideline values for more representative groups of individuals (relative to the entire population) than the current proposed guidelines. In this way population-attributable risk could be estimated. Factors that could be taken into account are age, sex, etc. The results should also be presented in disaggregated form to allow the effects of different sources of variation in risk to be compared.

Sensitivity analysis should examine the effects of assumptions made during risk assessment. These include independence of risks between different exposure occasions, and between different individuals in a family group. The sensitivity of the guideline values to the assumed standard deviation for $\log_{10}$ faecal streptococci counts could also be investigated. The value used is an average over several EC Member States. Values for individual Member States could be used instead. Finally, the sensitivity of the guideline values to the UK trial data could be explored by extending the risk assessment to use dose-response models from other studies, for example, Cabelli *et al.* (1982).

References

Balarajan R (1993) *Health Risks Associated with Bathing in the Sea: Results of Studies in Cleethorpes, Skegness, Instow and Westward Ho!*, Guildford, UK, Institute of Public Health, University of Surrey

Bandaranayake DR, Turner SJ, Mcbride GB, Lewis GD & Till DG (1995) *Health Effects of Bathing at Selected New Zealand Marine Beaches*, Wellington, New Zealand, Ministry of the Environment

Barnett V & O'Hagan A (1997) *Setting Environmental Standards — The Statistical Approach to Handling Uncertainty and Variation*, London, UK, Chapman & Hall

Cabelli VJ, Dufour AP, McCabe LJ & Levin MA (1982) Swimming-associated gastroenteritis and water quality. *Am J Epidemiol, 115,* 606–616

Cabelli VJ (1983) *Health Effects Criteria for Marine Recreational Waters* (EPA-600/1-80-031), Research Triangle Park NC, USA, US Environmental Protection Agency

Calderon RL, Mood EW & Dufour AP (1991) Health effects of swimmers and non-point sources of contaminated water. *Int J Environ Health Res, 1,* 21–31

Cheung WHS, Chang KCK, Hung RPS & Kleevens JWL (1990) Health effects of beach water pollution in Hong Kong. *Epidemiol Infect, 105,* 139–162

Corbett SJ, Rubin GL, Curry GK, Kleinbaum DG & The Sydney Beach Users Advisory Group (1993) The health effects of swimming at Sydney beaches. *Am J Public Health, 83,* 1701–1706

Covello VT & Merkhofer MW (1993) *Risk Assessment Methods — Approaches for Assessing Health and Environmental Risks*, New York NY, USA, Plenum Press

CREH (1996) *Relevance of Faecal Streptococci as Indicators of Pollution — A Report to DG XI of the Commission of the European Communities*, Leeds, UK, Centre for Research into Environment and Health, University of Leeds

Dalton CB, Mintz ED, Well JG, Bopp CA & Tauxe RV (1999) Outbreaks of enterotoxigenic *Eschericia coli* infection in American adults: A clinical and epidemiological approach. *Epidemiol Infect, 123,* 9–16

EPA (1976) *Quality Criteria for Water*, Washington DC, USA, US Environmental Protection Agency

EPA (1986) *Ambient Water Criteria for Water* (EPA 440/5-84-002), Washington DC, USA, US Environmental Protection Agency

Fattal B, Peleg-Olevsky E, Shuval HI & Agursky T (1987) The association between seawater pollution as measured by bacterial indicators and morbidity among bathers at Mediterranean bathing beaches in Israel. *Chemosphere, 16,* 565–570

Ferley JP, Zmirou D, Balducci F, Baleux B, Fera P, Larbaigt G, Jacq E, Moissonnier B, Blineau A & Boudot J (1989) Epidemiological significance of microbiological pollution criteria for river recreational waters. *Int J Epidemiol, 18,* 198–205

Figueras MJ, Polo F, Inza I & Guarro J (1997) Past, present and future perspectives of the EU Bathing Water Directive. *Mar Pollut Bull, 34,* 148–156

Fleisher JM (1990a) The effects of measurement error on previously reported mathematical relationship between indicator organism density and swimming-associated illness: A quantitative estimate of the resulting bias. *Int J Epidemiol, 19,* 1100–1106

Fleisher JM (1990b) Conducting recreational water quality surveys: Some problems and suggested remedies. *Mar Pollut Bull, 21,* 562–567

Fleisher JM (1991) A reanalysis of data supporting US federal bacteriological water quality criteria governing marine recreational waters. *Res J Water Pollut Control Fed, 63,* 259–265

Fleisher JM, Kay D, Salmon RL, Jones F, Wyer MD & Godfree AF (1996) Marine waters contaminated with domestic sewage: Nonenteric illnesses associated with bather exposure in the United Kingdom. *Am J Public Health, 86,* 1228–1234

Geldreich EE (1977) Bacterial populations and indicator concepts in faeces, sewage, stormwater and solid wastes. In: Berg G, ed, *Indicators of Virus in Water and Food*, Ann Arbor MI, USA, Ann Arbor Science, pp 51–97

Godfree AF, Kay D & Wyer MD (1997) Faecal streptococci as indicators of faecal contamination in water. *Soc Appl Bacteriol Symp Ser, 26,* 110S–119S

Haile RW, Witte JS, Gold M, Cressey R, McGee C, Millikan RC, Glasser A, Harawa N, Ervin C, Harmon P, Harper J, Dermand J, Alamillo J, Barrett K, Nides M & Wang G-Y (1999) The health effects of swimming in ocean water contaminated by storm drain runoff. *Epidemiology, 10,* 355–363

Jones F, Kay D, Wyer MD, Fleisher JM, Salmon R & Godfree AF (1993) *Final Report of the Controlled Cohort Investigations into the Health Effects of Bathing in Sewage Contaminated Coastal Waters,* Lampeter, UK, Centre for Research into Environment and Health, University of Wales

Jones K & Obiri-Danso K (1999) Non-compliance of beaches with the EU directives of bathing water quality: Evidence of non-point sources of pollution in Morecambe Bay. *J Appl Microbiol Symp Suppl, 85,* 101S–107S

Kay D, Fleisher JM, Salmon RL, Jones F, Wyer MD, Godfree AF, Zelenauch-Jacquotte Z & Shore R (1994) Predicting likelihood of gastroenteritis from sea bathing: Results from randomised exposure. *Lancet, 344,* 905–909

Kueh CSW, Tam T-Y, Lee T, Wong SL, Lloyd OL, Yu ITS, Wong TW, Tam JS & Bassett DCJ (1995) Epidemiological study of swimming-associated illnesses relating to bathing-beach water quality. *Water Sci Technol, 31,* 1–4

Mariño FJ, Moriñigo MA, Martinez-Manzanares E & Borrego JJ (1995) Microbiological-epidemiological study of selected marine beaches in Malaga (Spain). *Water Sci Technol, 31,* 5–9

MNHW (1992) *Guidelines for Canadian Recreational Water Quality,* Ottawa, Canada, Canadian Government Publishing Centre, Ministry of National Health and Welfare

Mujeriego R, Bravo JM & Feliu MT (1982) Recreation in coastal waters, public health implications. *Vlemes J Etud Pollut,* 585–594

NTAC (1968) *Water Quality Criteria*, Washington DC, USA, Federal Water Pollution Control, National Technical Advisory Committee

Nelson C & Williams AT (1997) Bathing water quality and health implications. In: *Proceedings of Water Pollution IV — Modelling, Measuring and Prediction, June 1997, Bled, Slovenia,* Southampton, UK, Computational Mechanics Publications, pp 175–183

Nurminen M, Nurminen T & Corvalán CF (1999) Methodologic issues in epidemiologic risk assessment. *Epidemiology, 10,* 585–593

Nuzzi R & Burhans R (1997) The use of enterococcus and coliform in characterising bathing-beach waters. *J Environ Health, 60,* 16–22

Olivieri VP (1982) Bacterial indicators of pollution. In: Pipes WO, ed, *Bacterial Indicators of Pollution*, Boca Raton Fl, USA, CRC Press, pp 21–41

PHLS (1959) Sewage contamination of coastal bathing waters in England and Wales: A bacteriological and epidemiological study. *J Hyg, 43,* pp 435–472

Pike EB (1991a) *Health Effects of Sea Bathing (ET 9511) Phase I — Pilot Studies at Langland Bay 1989* (DoE 2518-M(P)), Medmenham, UK, Water Research Centre Plc

Pike EB (1991b) *Health Effects of Sea Bathing (EM 9511) Phase II — Studies at Ramsgate and Moreton 1990* (DoE 2736-M), Medmenham, UK, Water Research Centre Plc

Pike EB (1994) *Health Effects of Sea Bathing (WMI 9021) — Phase III — Final Report to the Department of the Environment* (DoE 3412/2), Medmenham, UK, Water Research Centre Plc

Prüss A (1998) Review of epidemiological studies on health effects from exposure to recreational water. *Int J Epidemiol, 27,* 1–9

Richman DD, Whitley RJ & Hayden FG, eds (1997) *Clinical Virology*, Edinburgh, UK, Churchill Livingstone

Rothman KJ & Greenland S (1998) *Modern Epidemiology* (2nd Edition), Philadelphia PL, USA, Lippincott-Raven Publishers

Serrano E, Moreno B, Solaun M, Aurrekoetxea JJ & Ibarluzea J (1998) The influence of environmental factors on microbiological indicators of coastal water pollution. *Water Sci Technol, 38,* 195–199

Spraycar M, ed (1995) *Stedman's Medical Dictionary* (Twenty sixth Edition), Baltimore MD, USA, Williams & Wilkins

Van Dijk PAH, Lacey RF & Pike EB (1996) *Health Effects of Sea Bathing — Further Analysis of Data from UK Beach Surveys* (DoE 4126/3), Medmenham, UK, Water Resource Centre Plc

Wang G-Y (1998) An epidemiological study of possible adverse effects of swimming in Santa Monica Bay. In: *Proceedings of the California and World Ocean Conference, San Diego, California, March, 1997,* Reston VA, USA, American Society of Civil Engineers, pp 869–875

Wheeler JG, Sethi D, Cowden JM, Wall PG, Rodrigues LC, Tompkins DS, Hudson MJ & Roderick PJ (1999) Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. *Brit Med J,* 318, 1046–1059

WHO (1998) Guidelines for Safe Recreational-Water Environments: Coastal and Fresh-Waters (*Draft for Consultation).* Geneva, Switzerland, World Health Organization

Annex I: Confounders

The following variables recorded during the randomised controlled trials were identified by Kay *et al.* (1994, Table 1) as potential confounders of the relationship between gastroenteritis and water quality:

- age;
- sex;
- usual frequency of alcohol consumption;
- history of migraine headaches;
- history of stress or anxiety;
- frequency of diarrhoea;
- current use of prescription drugs;
- illnesses in the four weeks preceding exposure that lasted for more than 24 hours;
- use of prescription drugs in the four weeks preceding exposure;
- taking of laxatives in the four weeks preceding exposure;
- taking of other stomach remedies in the four weeks preceding exposure;
- consumption of mayonnaise in the period from three days before exposure to seven days afterwards;
- consumption of purchased sandwiches in the period from three days before exposure to seven days afterwards;
- consumption of chicken in the period from three days before exposure to seven days afterwards;
- consumption of eggs in the period from three days before exposure to seven days afterwards;
- consumption of hamburgers in the period from three days before exposure to seven days afterwards;
- consumption of hotdogs in the period from three days before exposure to seven days afterwards;

- consumption of raw milk in the period from three days before exposure to seven days afterwards;
- consumption of cold meat pies in the period from three days before exposure to seven days afterwards;
- consumption of seafood in the period from three days before exposure to seven days afterwards;
- consumption of alcohol in the seven days following exposure;
- illnesses in the household in the three weeks following exposure; and
- additional bathing in the period from three days before exposure to three weeks afterwards.

Pike (1994, Table 4.2) lists further variables on which data were collected during the randomised controlled trials. These are also potential confounders. They include:

- occupation;
- smoking habits;
- history of overseas residence;
- general leisure activities;
- consumption of ice-cream in the period from three days before exposure to seven days afterwards;
- consumption of salad in the period from three days before exposure to seven days afterwards;
- consumption of cold meats and pâté in the period from three days before exposure to seven days afterwards;
- consumption of any take-away foods in the period from three days before exposure to seven days afterwards;
- history of sunburn;
- history of motion sickness;
- awareness of beach maintenance in the UK;
- awareness of pollution; and
- awareness of news/media coverage of the trials.

Further variables on which information was collected during the randomised controlled trials are evident in the questionnaires used for the trials (see Pike, 1991a, b; Jones *et al.*, 1993). Potential confounders not listed above include:

- holidays/business trips in the UK and abroad in the period from four weeks before exposure to three weeks afterwards;
- views on cleanliness of UK beaches; and
- concern about the way in which UK beaches are maintained.

Bathers were closely supervised during the periods they spent in the sea (Jones *et al.*, 1993). The supervisors used diary sheets to record the following information relating to exposure:

- duration of exposure;
- depth of water at the bather's location (recorded at ten-minute intervals during exposure);
- type of activity in which the bather was engaged (paddling/wading, swimming, or full immersion; recorded at ten-minute intervals during exposure); and
- ingestion of water (recorded on leaving the water).

These variables are potential effect-modifiers in that they might modify the relationships between gastroenteritis and bacterial indicators of water quality.

Annex II: Studies relating gastrointestinal symptoms to indicators other than faecal streptococci, enterococci, *E. coli* and faecal coliforms, and studies relating other health outcomes to bacterial indicators of water quality

Table B.1 Studies relating gastrointestinal symptoms to total coliform and other indicators of water quality

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Cheung *et al.* (1990)	*Klebsiella* spp	GI	Non-bathers		Not reported	'Incidence rates' are swimming-associated symptom rates (incidence rates for bathers minus incidence rates for non-bathers).
			0-100		1.8	
			101-1000		5.1	
		HCGI	Non-bathers		Not reported	
			0-100		0.9	
			101-1000		2.5	
	Staphylococci	GI	Non-bathers		Not reported	
			0-1000		3.9	
			1001-3000		4.5	
		HCGI	Non-bathers		Not reported	
			0-1000		2.2	
			1001-3000		1.8	
	P. aeruginosa	GI	Non-bathers		Not reported	
			0-5		4.8	
			6-45		3.7	
		HCGI	Non-bathers		Not reported	
			0-5		2.2	
			6-45		1.7	
	Candida albicans	GI	Non-bathers		Not reported	
			0-6		3.3	
			7-20		4.4	
		HCGI	Non-bathers		Not reported	
			0-6		0.7	
			7-20		2.4	

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Cheung *et al.* (1990)	Total fungi	GI	Non-bathers		Not reported	
			0-140		5.5	
			141-770		3.3	
		HCGI	Non-bathers		Not reported	
			0-140		2.0	
			141-770		1.9	
Haile *et al.* (1999)	TC	HCGI	0-1000		32	No non-bathers.
			1000-10000	0.84 (0.62-1.14)	27	Relative risk was calculated
			10000+	0.74 (0.44-1.23)	22	in relation to lowest
		HCGI 2[b]	0-1000		10	exposure group (adjusted for
			1000-10000	0.89 (0.51-1.55)	8	age, sex, race, resident-or-not
			10000+	0.83 (0.32-2.12)	7	and concern about beach hazards).

GI, gastrointestinal illness; HCGI, highly credible GI; TC, total coliforms
[a]Organisms per 100 ml; [b]Vomiting and fever

Table B.2 Studies relating non-enteric outcomes to bacterial indicators of water quality

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Faecal streptococci and enterococci						
Cheung *et al.* (1990)	Ent	Ear	Non-bathers		Not reported	'Incidence rates' are swimming-associated symptom rates (incidence rates for bathers minus incidence rates for non-bathers).
			0-39		0.8	
			40-250		1.0	
		Eye	Non-bathers		Not reported	
			0-39		5.4	
			40-250		3.2	
		Skin	Non-bathers		Not reported	Details of pre-trial illness, bathing activity and food eaten not incorporated into any calculations/conclusions.
			0-39		2.2	
			40-250		11.2	
		Respiratory	Non-bathers		Not reported	
			0-39		8.9	
			40-250		11.4	
		Total	Non-bathers		Not reported	
			0-39		24.3	
			40-250		33.5	
	FS	Ear	Non-bathers		Not reported	
			0-55		0.9	
			56-290		0.9	
		Eye	Non-bathers		Not reported	
			0-55		6.5	
			56-290		3.1	
		Skin	Non-bathers		Not reported	
			0-55		3.1	
			56-290		9.4	
		Respiratory	Non-bathers		Not reported	
			0-55		11.1	
			56-290		10.1	
		Total	Non-bathers		Not reported	
			0-55		27.0	
			56-290		31.1	
Fleisher *et al.* (1996)	FS	Acute febrile respiratory illness	Non-bathers		30	Account taken of non-water related risk factors or possible confounders.
			0-14		40	
			15-27		58	
			28-50		16	
			51-158		76	

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Bandaranayake *et al.* (1995)	Ent	Respiratory	Non-bathers		23	Risk ratios adjusted for age group.
			0-1.5	0.77	24	
			1.5-3.75	1.18	37	
			3.75-13	1.40	42	
			13-232.25	2.91	71	
Faecal coliforms and *E. coli*						
Fattal *et al.* (1987)	*E. coli*	Respiratory			Not reported	Relative risk not associated with pollution.
		Skin disease			Not reported	Relative risk not associated with pollution.
		Ear infection			Not reported	No excess relative risk in bathers.
Cheung *et al.* (1990)	FC	Ear	Non-bathers		Not reported	'Incidence rates' are swimming-associated symptom rates (incidence rates for bathers minus incidence rates for non-bathers).
			0-410		0.9	
			411-3200		0.9	
		Eye	Non-bathers		Not reported	
			0-410		6.5	
			411-3200		3.1	
		Skin	Non-bathers		Not reported	Details of pre-trial illness, bathing activity and food eaten not incorporated into any calculations/conclusions
			0-410		3.1	
			411-3200		9.4	
		Respiratory	Non-bathers		Not reported	
			0-410		11.1	
			411-3200		10.1	
		Total	Non-bathers		Not reported	
			0-410		27.0	
			411-3200		31.1	

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Cheung *et al.* (1990)	*E. coli*	Ear	Non-bathers		Not reported	
			0-180		0.9	
			181-1800		0.9	
		Eye	Non-bathers		Not reported	
			0-180		5.2	
			181-1800		3.5	
		Skin	Non-bathers		Not reported	
			0-180		4.0	
			181-1800		10.0	
		Respiratory	Non-bathers		Not reported	
			0-180		10.5	
			181-1800		10.5	
		Total	Non-bathers		Not reported	
			0-180		27.8	
			181-1800		31.8	
Fleisher *et al.* (1996)	FC	Ear ailments	Non-bathers		28	Account taken of non-water related risk factors and possible confounders.
			0-40		67	
			41-79		62	
			80-133		59	
			134-661		142	
Kueh *et al.* (1995)	*E. coli*	Respiratory	Bathers (B1)		10.4	Symptom rates (corrected for non-bather rates) given for two beaches having seasonal geometric means of 13 *E. coli* per 100 ml (B1); and 237 *E coli* per 100 ml (B2). No account taken of confounders.
			Bathers (B2)		10.2	
		HC respiratory	Bathers (B1)		1.5	
			Bathers (B2)		1.1	
		Eye	Bathers (B1)		7.3	
			Bathers (B2)		8.5	
		Skin	Bathers (B1)		14.8	
			Bathers (B2)		14.0	
		Total illness	Bathers (B1)		39.2	
			Bathers (B2)		42.5	
Bandaranayake *et al.* (1995)	FC	Respiratory	0.375-3.5	1.68		Risk ratios adjusted for age group.
			3.5-8	0.82		
			8-35	1.76		
			35-724	1.41		

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Bandaranayake *et al.* (1995)	*E. coli*	Respiratory	0-3 3-6 6-28.25 28.25-552	1.69 0.74 1.99 1.22		
Cheung *et al.* (1990)	*Klebsiella* ssp	Ear	Non-bathers 0-100 101-1000		Not reported 0.9 0.9	'Incidence rates' are swimming-associated symptom rates (incidence rates for bathers minus incidence rates for non-bathers).
		Eye	Non-bathers 0-100 101-1000		Not reported 6.5 3.1	
		Skin	Non-bathers 0-100 101-1000		Not reported 3.1 9.4	
		Respiratory	Non-bathers 0-100 101-1000		Not reported 11.1 10.1	
		Total	Non-bathers 0-100 101-1000		Not reported 27.0 31.1	
	Staphylococci	Ear	Non-bathers 0-1000 1001-3000		Not reported 0.8 1.0	
		Eye	Non-bathers 0-1000 1001-3000		Not reported 5.3 3.3	
		Skin	Non-bathers 0-1000 1001-3000		Not reported 5.5 9.1	
		Respiratory	Non-bathers 0-1000 1001-3000		Not reported 5.8 14.0	
		Total	Non-bathers 0-1000 1001-3000		Not reported 25.7 33.3	

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Cheung *et al.* (1990)	*P. aeruginosa*	Ear	Non-bathers		Not reported	
			0-5		1.1	
			6-45		0.8	
		Eye	Non-bathers		Not reported	
			0-5		3.4	
			6-45		4.9	
		Skin	Non-bathers		Not reported	
			0-5		6.6	
			6-45		8.8	
		Respiratory	Non-bathers		Not reported	
			0-5		10.8	
			6-45		10.4	
		Total	Non-bathers		Not reported	
			0-5		30.9	
			6-45		29.7	
	C. albicans	Ear	Non-bathers		Not reported	
			0-6		0.8	
			7-20		1.0	
		Eye	Non-bathers		Not reported	
			0-6		4.4	
			7-20		4.0	
		Skin	Non-bathers		Not reported	
			0-6		6.5	
			7-20		8.0	
		Respiratory	Non-bathers		Not reported	
			0-6		14.5	
			7-20		8.7	
		Total	Non-bathers		Not reported	
			0-6		31.5	
			7-20		29.1	

Reference	Indicator	Health outcome	Exposure category[a]	Relative risk (95% confidence interval)	Incidence rate (per 1000 individuals)	Comments
Cheung *et al.* (1990)	Total fungi	Ear	Non-bathers		Not reported	
			0-140		1.2	
			141-770		0.7	
		Eye	Non-bathers		Not reported	
			0-140		4.2	
			141-770		4.0	
		Skin	Non-bathers		Not reported	
			0-140		7.3	
			141-770		7.9	
		Respiratory	Non-bathers		Not reported	
			0-140		16.6	
			141-770		6.4	
		Total	Non-bathers		Not reported	
			0-140		35.0	
			141-770		26.6	

[a] Organisms per 100 ml