Setting microbiological water quality standards for sea bathing – a critical evaluation

M.A. Mugglestone, E.D. Stutt and L. Rushton
MRC Institute for Environment and Health, University of Leicester, 94 Regent Road, Leicester, LE1 7DD, UK
(E-mail: mam14@le.ac.uk)

Abstract The World Health Organization (WHO) recently produced draft guidelines for safe recreational water environments. The microbiological standards proposed in the guidelines are expected to over-estimate the degree of water quality required to provide given levels of public health protection. The WHO standards were obtained by means of a risk assessment which featured a dose-response model derived from a series of randomised controlled trials. The trials have many strengths but biases and problems with statistical analysis are likely to have led to over-estimation of the risks from bathing in the dose-response model. In addition, the WHO risk assessment failed to consider the effects of uncertainty and variability in risk estimates and sensitivity to model assumptions. Improved standards could be obtained by extending the risk assessment to examine these effects and by incorporating a suitably revised dose-response model.

Keywords Bacterial indicators; environmental epidemiology; epidemiological randomised controlled trials; faecal streptococci; gastroenteritis; pollution; recreational water; swimming-associated illness

Introduction
In the 1970s, the European Community adopted the Bathing Water Directive (76/160/EEC) that sets, inter alia, guideline and mandatory standards for the microbiological quality of marine and fresh bathing waters. In the 1990s, the European Commission adopted a proposal to review the Directive. As part of the review, the Commission and EU Member States are considering draft guidelines for safe recreational water environments produced by the World Health Organization (WHO). Chapter 4 of the WHO guidelines (WHO, 1998) is most relevant to proposals to revise bathing water quality standards since it deals with microbiological aspects of water quality: The chapter includes an assessment of the epidemiological evidence for associations between adverse health effects and bathing water quality and a risk assessment which is used to calculate microbiological standards intended to provide given levels of public health protection. In order to determine the relevance of the WHO guidelines to the issue of standard setting, the UK Department of the Environment, Transport and the Regions (DETR) commissioned the Institute for Environment and Health to provide a critical evaluation of WHO (1998, Chapter 4). The evaluation was restricted to the adverse health effects of sea bathing, although WHO (1998, Chapter 4) refers to both marine and fresh water environments. Full details of the evaluation are presented in Mugglestone et al. (2000).

This paper summarises the main findings: (a) a review of the epidemiological literature relating to adverse health effects of sea bathing and a discussion of the weight given to the epidemiological evidence in WHO (1998, Chapter 4), (b) an assessment of the design, conduct and statistical analysis of a series of randomised controlled trials which feature prominently in the WHO risk assessment, (c) an evaluation of the WHO risk assessment itself, (d) recommendations for modifications to the risk assessment that could eliminate some of the problems identified and (e) a discussion of issues that should be considered in future epidemiological studies of adverse health effects of sea bathing.
Review of the epidemiological evidence for adverse health effects of sea bathing

Introduction

WHO (1998, Chapter 4) contains a review (also published as Prüss, 1998) of the epidemiological literature relating to adverse health effects of sea bathing. In the interests of brevity we have chosen to refer only to WHO (1998, Chapter 4) in this paper and have examined the literature for studies completed after Prüss (1998). The articles reviewed were the marine studies cited in WHO (1998, Table 4.1) plus Nelson and Williams (1997). We could not obtain copies of CSIR (1995) and UNEP (1991) cited in WHO (1998, Chapter 4) and used Haile et al. (1999) instead of Haile et al. (1996).

Summary of general issues relating to studies of adverse health effects of sea bathing

In most studies, the potential of bathing water to cause adverse health effects is assessed by enumerating indicator microorganisms in waters as proxies for pathogens associated with faecal pollution. Commonly used indicators include total and faecal coliforms, for which mandatory and guideline standards are specified in the Directive, and faecal streptococci (used interchangeably with the term enterococci) for which guideline standards are specified. Indicator counts vary according to the temporal and spatial proximity of sources of pollution and according to environmental conditions. Water samples intended to provide measurements of exposure to pollution should, therefore, be collected at the time and location of bathing but seasonal or study-period averages are sometimes used. Membrane filtration is most often used to estimate indicator organism density. Ideally one would want to make replicate determinations to avoid imprecise estimation of exposure but this is rarely done. Quality control analysis is sometimes undertaken on a subset of samples.

Most studies address mild to moderate self-limiting ailments (gastroenteritis, eye infections, skin complaints, ear, nose and throat infections, and respiratory illnesses). There is a tendency to rely on self-reported symptoms because the pathogens that cause illnesses may be unknown or impossible to isolate from clinical samples. Direct comparison of studies is complicated by the use of different criteria to define similar adverse health effects. All but a few investigations are conducted as cohort studies where healthy subjects are recruited at a beach and asked about recent bathing activity. A few days later, subjects are re-contacted and asked about their state of health. A different type of design is a randomised controlled trial in which subjects are randomly assigned to bathing and non-bathing groups. To date, this design has been used only once although results for different health outcomes have been reported in separate publications (Kay et al., 1994; Fleisher et al., 1996). Trials should yield more precise estimates of risks of experiencing adverse health effects than cohort studies because they are effectively experiments in which the investigators have some control over the degree, duration and timing of exposure. Given appropriate data, the analysis of cohort studies and randomised controlled trials can adjust for the effects of confounders (e.g. age) that are associated with adverse health effects in the absence of exposure and are also associated with exposure. Analysis of both types of study can include the fitting of dose-response models to quantify relationships between health effects and exposure.

Evaluation of the WHO literature review

WHO (1998, Table 4.4) summarises (and in some cases recalculates) unadjusted relative risks and confidence intervals from various epidemiological studies of the adverse health effects of sea bathing. Our comments are based on incidence rates and relative risks reported in the original articles with adjustment for confounders where reported. The baseline used in the calculation of relative risks in these articles is sometimes the risk to non-bathers and sometimes the risk to bathers in the lowest exposure category. WHO (1998, Chapter 4) concludes that gastroenteritis is the most frequently studied adverse health effect and that
faecal streptococci and enterococci are most frequently associated with adverse health effects. These conclusions appear to be reasonable. Incidence rates of gastroenteritis are generally higher in bathers than non-bathers but there is not always a consistent increase in incidence with decreasing water quality. In fact, only two studies (Cabelli et al., 1982; Kay et al., 1994) report dose-response relationships between gastroenteritis incidence and water quality. WHO (1998, Chapter 4) gives the greatest weight to results reported by Kay et al. (1994) and uses their dose-response model to formulate a risk assessment. This appears to be a logical choice because that report combines: (a) randomised controlled trials, (b) a dose-response relationship linking gastroenteritis incidence to faecal streptococci counts measured at the time and location of exposure, and (c) recording of potential confounding variables. In order to appreciate the implications of the decision to rely on this dose-response model (Kay et al., 1994) it is useful to examine the design and conduct of the trials on which the model is based, the statistical analysis reported by Kay et al. (1994) and some differences between the latter’s results and those of other studies relating gastroenteritis to indicators of faecal pollution.

The UK trials

Introduction

The dose-response model of Kay et al. (1994) was based on data from randomised controlled trials conducted at four bathing beaches in the UK during the period 1989-92. The trials, which were commissioned by the UK government, were designed to examine the risks of contracting minor illnesses from bathing in sewage-contaminated water. Ten prospective cohort studies were also conducted at UK bathing beaches as part of the same research programme. The results of both series of studies have been reported in various publications (Pike 1991a,b; Jones et al., 1993; Balarajan, 1993; Pike, 1994; Kay et al., 1994; Fleisher et al., 1996; van Dijk et al., 1996). All of these were examined as part of our review. The UK trials have many strengths (Table 1).

Potential for bias

Despite the many strengths of the trials, there was some potential for bias as indicated in Table 2. Subjects were recruited in advance of the trials by means of publicity which included reports in the media, subject information sheets identifying stomach infections as potential adverse health effects and displays showing information from earlier trials (Pike, 1991a, b; Jones et al., 1993).

Subjects were, therefore, alerted to likely adverse health effects. The subjects were all healthy adult volunteers. This restriction, which was required for ethical approval of the trials, meant that individuals with weakened immune systems were excluded. Self-reporting

Table 1 Strengths of the UK trials

- Subjects randomised to bathing and non-bathing groups
- Duration of exposure partially controlled (subjects spent at least 10 min in the water)
- Degree of exposure partially controlled (subjects immersed their heads at least 3×)
- Exposure measured using recognised indicators (total and faecal coliforms, faecal streptococci, total staphylococci and Pseudomonas aeruginosa)
- Microbiological samples obtained at the time and location of exposure
- Some quality control analysis of microbiological samples undertaken
- Data on potential confounders collected
- Data on adverse health effects collected by medical examination as well as by self-reporting of symptoms
- Gastroenteritis treated as the primary adverse health effect
of symptoms was used in interviews and postal questionnaires. A tendency to report mild symptoms could have led to over-reporting in both bathers and non-bathers. Differential recall between bathers and non-bathers could have led to over-estimation of the relative risk of bathing (bathers compared to non-bathers).

Biases introduced through recruitment methods, restriction to healthy adults and self-reporting of symptoms stem from the design and conduct of the trials. These forms of bias cannot be eliminated retrospectively, nor can they be quantified, although interpretation of the study results should attempt to take them into account. Indeed, it would be difficult to avoid these forms of bias even if new studies were to be conducted. Epidemiological trials are, therefore, less precise than randomised controlled drug trials in which subjects (and often investigators) are blind to the treatments they receive and in which far greater control over the behaviour of subjects can be exercised. The only source of bias highlighted by Kay et al. (1994) and WHO (1998, Chapter 4) was the restriction to healthy adult volunteers. This is the one source for which under-estimation of risks to bathers would be expected to occur. WHO (1998, Chapter 4) acknowledges the potential for media and publicity to induce bias but does not discuss the extent to which this could have affected the UK trials.

The other sources of bias in Table 2 relate to the way in which the data were analysed. Kay et al. (1994) analysed incidences of “objective” gastroenteritis occurring up to 21d after exposure. Gastroenteritis can be caused by bacteria, parasites and viruses with varying incubation periods of <14 d. A recent study (Wheeler et al., 1999) identified, where possible, the aetiological agents responsible for gastroenteritis in the general population. The majority of these agents have incubation periods of <7 d (Spraycar, 1995; Richman et al., 1997). Only 34.3% of the cases reported by Kay et al. (1994) developed within 7 d of exposure suggesting that substantial over-estimation of the risks of experiencing gastroenteritis due to controlled exposure might have occurred. Their definition of objective gastroenteritis was “any case of vomiting or diarrhoea plus any case of either indigestion or nausea accompanied by fever”. This was slightly less stringent than the definition put forward by Cabelli et al. (1982) of “highly credible” gastroenteritis that approximated clinically defined viral gastroenteritis, and this could have led to over-estimation of the risks of experiencing gastroenteritis. Kay et al. (1994) reported strikingly higher incidence rates of gastroenteritis in bathers and non-bathers than do other studies. The only results that approach those obtained by Kay et al. (1994) were reported by Fattal et al. (1987) for 0–4 year olds (in whom the incidence rate of gastroenteritis is generally high). The extent to which differences between the rates reported in Kay et al. (1994) and in other studies might reflect bias linked to the 3-week observation period and the definition of gastroenteritis could be investigated by further statistical analysis.

Further issues related to the statistical analysis

Some other aspects of statistical analysis in Kay et al. (1994) give rise to concern (Table 3). They criticised other investigators for not measuring or adequately controlling for

---

**Table 2** Bias in the UK trials and expected effects on estimated risks of experiencing gastroenteritis

<table>
<thead>
<tr>
<th>Source of bias</th>
<th>Risk to bathers (A)</th>
<th>Risk to non-bathers (B)</th>
<th>Relative risk (A/B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recruitment methods</td>
<td>↑</td>
<td>↓</td>
<td>↑</td>
</tr>
<tr>
<td>Restriction to healthy adults</td>
<td>↓</td>
<td>↓</td>
<td>↓↑</td>
</tr>
<tr>
<td>Self-reporting of symptoms</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>Length of observation period</td>
<td>↑</td>
<td>↑</td>
<td>↓</td>
</tr>
<tr>
<td>Definition of gastroenteritis</td>
<td>↑</td>
<td>↑</td>
<td>=</td>
</tr>
</tbody>
</table>

↓ tendency to under-estimate; ↑ tendency to over-estimate; = estimate unaffected
confounders but they did not investigate the effects of all potential confounders in their own study. They also failed to allow for confounders that were not significant individually but might have had appreciable effects in conjunction with other confounders. They implied that the final dose-response model adjusted for confounders but in fact the only explanatory variable in the model was the faecal streptococci count (Jones et al., 1993).

This is worrying because earlier analyses showed that the variable most strongly associated with incidence of gastroenteritis was the presence of gastroenteritis in the household preceding the subject’s own symptoms (Table 6 in Kay et al., 1994). The rate of increase of incidence of gastroenteritis with decreasing water quality (i.e. the slope of the dose-response curve) was greater in Kay et al. (1994) than in other studies. This could be due to lack of control for confounders. Their study included subjects who reported additional bathing activity during the period from 3 d before exposure to the end of the observation period whereas such subjects were excluded from analysis in other studies (e.g. Cabelli et al., 1982).

Kay et al. (1994) estimated the relationship between gastroenteritis incidence and water quality using data for bathers only. There are advantages and disadvantages to this approach. The main advantage is that, although subjects in the UK trials were not blind to the results of randomisation, bathers were blind to the quantity of faecal pollution in the water in which they bathed. Estimates of the relative risks of bathing in mildly and more heavily polluted waters should, therefore, be more reliable than comparisons between bathers and non-bathers. One disadvantage of excluding non-bathers is that reduction of the sample size from 1,112 to 507 would have reduced the precision of estimated risks of experiencing gastroenteritis. Another disadvantage is that the effect of contact with seawater itself cannot be investigated. Water-related activity (regardless of water quality) had statistically significant effects on incidence rates of gastroenteritis (and other adverse health effects) in the UK cohort studies (van Dijk et al., 1996). In early analyses, faecal streptococci measurements were represented by categorical variables with each category corresponding to a range of counts (Kay et al., 1994). Assuming that the desired end-product was a dose-response relationship linking incidence of gastroenteritis to a microbiological indicator of faecal pollution, it would have been preferable to have treated the count as a continuous explanatory variable. The counts were treated as a continuous variable in the final dose-response model obtained using logistic regression but a square-root transformation was applied to them on the grounds that this was appropriate given their statistical distribution. In fact, the distribution of the counts is immaterial when they are used as an explanatory variable in a regression model. The most appropriate transformation to use should have been determined by examining the fit of models corresponding to different transformations (e.g. using the counts themselves or log_{10} counts).

The final dose-response model (Kay et al., 1994) actually consisted of two logistic regression models, one for bathers exposed to fewer than 32 faecal streptococci/100 mL, the other for bathers exposed to at least 32 faecal streptococci/100 mL. The cut-off was chosen because it was the median of the exposure category that indicated no excess risk to

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Issues that could be clarified by further statistical analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Effects of confounding factors</td>
<td></td>
</tr>
<tr>
<td>• Effect of including data for non-bathers</td>
<td></td>
</tr>
<tr>
<td>• Selection of the most appropriate transformation of faecal streptococci counts to use as a continuous explanatory variable</td>
<td></td>
</tr>
<tr>
<td>• Extent of evidence for a threshold of risk</td>
<td></td>
</tr>
<tr>
<td>• Extent of uncertainty in the fitted model (measured by standard errors/confidence intervals)</td>
<td></td>
</tr>
</tbody>
</table>
bathers in previous analyses. The model for bathers exposed to fewer than 32 faecal streptococci/100 mL showed no statistically significant increase in risk with decreasing water quality. It was suggested that this provided evidence for a threshold of risk at 32 faecal streptococci/100 mL but this is a subjective method for determining a threshold. In any case, a statistically significant increase in risk is not necessarily the same as a clinically significant increase. No estimates of uncertainty (standard errors or confidence intervals) were reported by Kay et al. (1994) for their final dose-response model. Although trials are expected to be more precise than cohort studies, using a relatively small group of subjects could result in large standard errors and wide confidence intervals. Despite the limitations discussed above, the UK trials remain the only randomised controlled trials of adverse health effects of sea bathing. The degree and duration of exposure are expected to have been estimated more precisely than in other studies. Concerns raised in connection with the assessment of adverse health effects, and particularly with the dose-response model could be addressed by further statistical analysis. A suitably revised dose-response model would be appropriate for use in risk assessment.

The WHO risk assessment

The main elements of the WHO (1998, Chapter 4) risk assessment are shown in Table 4. The approach is based on: (a) characterisation of exposure, (b) characterisation of adverse health effects and (c) estimation of disease burden using a dose-response model that links health effects to an indicator of exposure and a probability distribution for indicator counts. These are widely accepted as being desirable elements in risk assessments relating adverse health effects to environmental exposures (Covello and Merkhofer, 1993; Barnett and O’Hagan, 1997; Nurminen et al., 1999; Haas, 2000).

Certain other important issues are not addressed. The first is the characterisation of the degree of uncertainty in risk estimates. Uncertainty arises because risks of experiencing gastroenteritis are estimated using a finite sample of subjects rather than by examining the entire population of would-be bathers. Uncertainty could be quantified by attaching standard errors and confidence intervals to the proposed guideline standards. Two major sources of uncertainty affecting the WHO risk assessment are the precision of the estimated standard deviation of log_{10} faecal streptococci counts and the precision of the dose-response model (Kay et al., 1994). All sources of variation in risk estimates should also be characterised. Potential sources of variation include the exposure measurement itself and the effects of confounders (e.g. age, sex) and other variables (e.g. geographical location/trial site). It is advisable to present the results of risk assessment at a disaggregated level (e.g. as separate

Table 4 Main elements of the WHO risk assessment

- Characterisation of exposure using faecal streptococci as an indicator of water quality
- Characterisation of adverse health effects in terms of gastroenteritis
- Adoption of a dose-response model (Kay et al., 1994) linking gastroenteritis incidence to faecal streptococci counts
- 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 less than or equal to 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 of the target value
estimates for different age groups and sexes). This could not have been attempted in the risk assessment framework presented by WHO (1998, Chapter 4) because the dose-response model (Kay et al., 1994) did not allow for the effects of confounders.

Finally, sensitivity analysis should be conducted to examine the effects of model assumptions on risk estimates. Two major assumptions made in the interpretation of the WHO (1998, Chapter 4) proposed guideline standards are that the risks experienced by a single bather on different exposure occasions are independent and that the risks experienced by different bathers in a family group are independent. The effects of relaxing these assumptions could be examined as part of sensitivity analysis. The effect of relying on the Kay et al. (1994) dose-response model rather than, say, that of Cabelli et al. (1982) could also be investigated. The problems we have highlighted with the statistical analysis of Kay et al. (1994) will be compounded when the dose-response model is used to calculate guideline standards. For this reason, the present guideline standards are expected to overestimate the degree of water quality required to provide a given level of health protection.

**Recommendations**

**Further analysis of the UK trial data**

Some of the concerns raised in relation to the dose-response model of Kay et al. (1994) could be addressed by further statistical analysis. We recommend the use of logistic regression models to examine incidence of gastroenteritis in bathers and non-bathers in relation to faecal streptococci counts. The response variable (incidence of gastroenteritis) should be defined to reflect the severity of symptoms used in other studies (e.g. Cabelli et al., 1982) and the expected incubation period (e.g. symptoms recorded within 0–7, 8–14 and 15–21 d after exposure). Models for different incubation periods should be compared with each other and with results from other studies.

The explanatory variables to be used in the models should include potential confounders (particularly, age, sex, consumption of foods associated with gastroenteritis, illness in the household between exposure and the end of symptom recording, additional bathing in the period from 3d before exposure to the end of symptom recording, and holidays/business trips in the UK and abroad in the period from four weeks before exposure to the end of symptom recording). Representation of two or more confounders as a single composite variable should be avoided. Geographical location (trial site) should be included as a factor or random effect. The remaining explanatory variables will be those related to exposure. An indicator variable representing exposure status (bather/non-bather) will allow the effect of contact with seawater to be estimated. The quality of the water to which bathers were exposed should be represented by a continuous variable based on faecal streptococci counts. The interaction between this term and the indicator variable representing exposure status will allow the effect of water quality to be estimated. The effects of other aspects of exposure could be investigated using data collected during the trials (for example, the duration of exposure and whether or not water was ingested).

Terms that are not measurements of exposure should be fitted first and an analysis of deviance should be performed to compare the strengths of the effects of different confounders and exposure variables. Standard errors or confidence intervals for risk estimates should be reported. The analysis should include an assessment of the form of the continuous variable representing faecal streptococci counts that gives the best fit. Analysis according to the specifications listed above will eliminate bias due to the length of the observation period and the definition of gastroenteritis. Sources of bias that cannot be quantified by re-analysis are recruitment methods, restriction to healthy adults and self-reporting of symptoms. Interpretation of the results should attempt to incorporate qualitative statements regarding the expected effects of these sources of bias.
Improvements to the WHO risk assessment

The WHO (1998, Chapter 4) risk assessment could be improved by incorporating a revised dose-response model based on the UK trial data. Extensions to the risk assessment framework to allow for uncertainty and variability in risk estimates and to examine the sensitivity of the results to model assumptions would also be beneficial. These extensions could be achieved by the use of Monte Carlo simulation (e.g. Covello and Merkhofer, 1993; Barnett and O’Hagan, 1997; Nurminen et al., 1999). Analysis by any other appropriate statistical method, such as Bayesian inference, would be equally acceptable. The analysis of uncertainty in risk estimates should address the uncertainty attached to the revised dose-response model and that due to the unknown distribution of $\log_{10}$ faecal streptococci counts. Variations in risk due to confounders and geographical locations should be quantified. Guideline values appropriate to different sectors of the population (depending on age, sex, etc.) should be presented, and population attributable risk should be estimated. The sensitivity of the guideline values to the assumption of independent risks, the estimate of the standard deviation of $\log_{10}$ faecal streptococci counts and the reliance on the UK trial data should be examined. Our review has not addressed the clinical significance and public health impact of the results of the UK trials and of the WHO (1998, Chapter 4) risk assessment. Assessment of any revised guideline standards should also include consideration of these issues.

Design, conduct and statistical analysis of future epidemiological studies

Our review has highlighted advantages and disadvantages of investigating adverse health effects of sea bathing using randomised controlled trials as opposed to cohort studies. The main advantage of randomised controlled trials is that they can be expected to yield more precise estimates of the risks of experiencing adverse health effects than cohort studies. This is due to the fact that subjects are randomised to bathing and non-bathing groups and that the degree and duration of exposure are controlled to some extent. However, randomised controlled trials of sea bathing are less precise than randomised controlled trials of medical interventions because it is difficult to control exposure before and after the trial day. Without having standard errors for the dose-response model of Kay et al. (1994) it is difficult to quantify the relative precision of randomised controlled trials and cohort studies.

One disadvantage of randomised controlled trials of sea bathing is the possibility of introducing bias into the assessment of adverse health effects because of inherent difficulties in blinding subjects to their exposure status. Other disadvantages arise from the need to obtain ethical approval for controlled trials. For example, the demographic characteristics of the study population may differ from those of the entire population of potential bathers, and the levels of pollution to which subjects can be exposed may be restricted by requiring studies to be conducted at beaches that meet current standards. These constraints may limit the extent to which randomised controlled trials can yield risk estimates for vulnerable groups of individuals and highly polluted waters. Finally, the costs of administering randomised controlled trials may be regarded as prohibitive. The USEPA has already announced that its future studies of the adverse health effects of recreational bathing are likely to be conducted as cohort studies (EPA, 1999).

Conclusions

This review highlighted the complexity of the statistical analysis needed for randomised controlled trials and cohort studies. The importance of: (i) adjusting for confounding factors, (ii) reporting standard errors or confidence intervals for fitted models and (iii) acknowledging potential sources of bias has already been discussed. Perhaps the best way of dealing with a large collection of potential confounders is to use a recognised model.
selection procedure such as forward selection (e.g. Rothman and Greenland, 1998) to identify those factors that are associated with adverse health effects at some low level of statistical significance (for example 10%) before examining the effects of variables that represent exposure measurements. Other issues relevant to the statistical analysis of randomised controlled trials and cohort studies include the dangers of attempting to assess associations between several adverse health effects and several different indicators of faecal pollution. Applying statistical tests at, say, the 5% level of significance means that one in 20 tests can be expected to yield a significant association just by chance. Repeated statistical testing can be minimised by identifying a primary adverse health effect and the faecal indicators most plausibly related to it at the design stage. Another problem associated with multiple testing is model-selection bias: the model that fits a given data set most closely will necessarily have smaller standard errors and confidence intervals than other models considered. Adjustments for this type of bias can be made using bootstrap techniques (e.g. Rothman and Greenland, 1998). Finally, it may be worth considering errors-in-variables models to allow for imprecision in indicator enumeration and multilevel models to allow for correlation between responses for subjects from the same family (Rothman and Greenland, 1998).

Acknowledgements
This work was funded by the DETR and peer-reviewed by members of the Public Health Laboratory Service Advisory Committee on Water and the Environment whose comments are taken into account in this paper. The DETR has recently formed a steering group to oversee further analysis of the UK trial data along the lines proposed here.

References


