Ministry for the Environment
Manatū Mō Te Taiao

MINISTRY OF HEALTH
MANATŪ HAUORA

Freshwater Microbiology Research Programme Report

Pathogen Occurrence and Human Health Risk Assessment Analysis

NOVEMBER 2002

Freshwater Microbiology Research Programme

Pathogen Occurrence and Human Health Risk Assessment Analysis

Graham McBride, NIWA, Hamilton
Desmond Till, Consultant, Wellington
Dr Terry Ryan, Animal Health Board, Hamilton
Andrew Ball, ESR, Christchurch
Dr Gillian Lewis, University of Auckland
Dr Stephen Palmer, Medical Officer of Health, Wellington
Professor Philip Weinstein, Wellington School of Medicine

27 November 2002

EXECUTIVE SUMMARY

In 1995 a survey of opinions from regional councils revealed the need for freshwater guidelines to help in the management of fresh waters used for recreation. This request was supported by the Ministries for the Environment, Health, and Agriculture and Forestry. The study for a "Freshwater Microbiological Research Programme" was funded by the Ministry of Research Science and Technology in 1997.

A Management Group was set up to guide and oversee the design and implementation of the programme, using a regional council representative, government department staff, a consultant from the private sector, and two Crown Research Institutes.

The ten indicators and pathogens used were: *E. coli*, *Clostridium perfringens* spores, FRNA bacteriophage, somatic coliphage, enteroviruses, adenoviruses, *Cryptosporidium* oocysts, *Giardia* cysts, Salmonellae and *Campylobacter*. A preliminary study at 3 sites was used to test sampling and laboratory methods.

Following the preliminary study, 25 sites distributed throughout New Zealand were selected on the basis of representing five different categories of predominant environmental impact (land use, waste discharge, and waterfowl activity). These are **B** (birds), **D** (dairy farming), **F** (forestry/undeveloped), **M** (municipal), and **S** (sheep/pastoral). In addition, five of the sites were also source waters for treated drinking-water supplies.

The main study has measured these ten variables fortnightly at the 25 sites for 15 months, from December 1998 to February 2000 (with a one-month gap). Twelve regional councils carried out the sampling, with four accredited laboratories performing the tests. Ancillary weather and water quality data were also recorded at the time of each sampling along with any noted unusual activities (e.g., increased waterfowl and bird densities, cattle in stream). A detailed catchment assessment (sanitary survey) was conducted for each site, covering a minimum distance of two kilometres upstream for those sites on rivers.

The results have formed the basis of the risk assessment presented herein. Associated studies have also been completed on stock-water drinking patterns and farm trough surveys, and some analysis of current Canadian studies on the effects of water quality on cattle.

The main outcomes of the risk assessment were that:

- Of the pathogens assessed in this study, *Campylobacter* and human adenoviruses are the pathogens most likely to cause human waterborne illness to recreational freshwater users.
- Using data from all sites, an estimated 4% of notified campylobacteriosis in New Zealand could be attributable to water contact recreation.
- The critical value for *E. coli* as an indicator of increased *Campylobacter* infection is in the range of 200-500 *E. coli* per 100mL.
- Infection risks of other pathogens examined have not been able to be related to *E. coli* concentrations in fresh waters.

Other findings include:

The most commonly used faecal indicator for fresh waters, *E. coli,* was detected in 99% of all samples, with somatic coliphage being detected most of the time (89%). Both tended to be highest in summer/autumn. *Clostridium perfringens* spores and FRNA phage were detected in only about half of all samples (58% and 52% respectively). They tended to be elevated in winter.

The *Campylobacter* detection rate was 60%; furthermore 43 (i.e., 5%) of all samples were above the test's detection limit. They were highest in late summer-early autumn. *C. jejuni* was the most frequent thermotolerant species identified, being present in at least 48% of the positive samples.

Each virus group was detected in about one third of all samples, though they were strongly dissociated (i.e., if adenovirus was detected in a sample enterovirus was seldom detected, and vice versa; a virus was detected in 59% of samples). No clear temporal pattern was obvious.

The *Salmonella* detection rate was low (10% of samples), and was highest in August and associated with outbreaks in sheep and associated human cases.

Giardia cysts and *Cryptosporidium* oocysts were detected rather infrequently (8% and 5% respectively) and at low concentrations in all catchment types, especially compared with the preliminary survey.

Correlations between data collected within the bathing season are very similar to those obtained using all the data. Correlations between indicators and pathogens were generally low. Somatic coliphage was well correlated with *E. coli* and there were moderate correlations between somatic coliphage and *Campylobacter*, between somatic coliphage and FRNA phage, and also between *E. coli* and *Campylobacter*. However, correlations are influenced by catchment types; in the catchment groups with higher levels of *Campylobacter* (**S** and **M**) its correlation with *E. coli* is stronger. The largest proportion of high *Campylobacter* values occurred in the **S** catchments and the greatest spread of values occurred in the **M** catchments.

Bird catchments were the most contaminated, across nearly all microorganisms.

Dairying catchments were often the second-most contaminated, but not for *Campylobacter* nor for adenoviruses.

The Municipal (**M**) and Forestry/Undeveloped (**F**) catchments were generally the least contaminated.

The overall pattern of distribution of *Campylobacter* species was similar between catchment types, except that the Sheep (**S**) catchments contained elevated levels of *C. lari* (33% of positive samples, versus 14% average for all other catchments).

Turbidity and catchment type are important explicatory variables for indicators and pathogens in fresh waters.

Table of Contents

1 INTRODUCTION

1.1 Origins of the Programme

In 1995 a survey of opinions from Regional Councils (Pyle, 1995) revealed that an important need for aquatic resources management information concerned microbiological water quality as it affects recreational use of freshwaters (i.e., swimming, water-skiing and wind-surfing). Their opinions were fuelled by the widespread incidence of elevated levels of microbiological indicators of faecal material, particularly—but not only—in areas impacted by agricultural activities (Smith *et al*., 1993). Supporting that view was a parallel understanding of the inevitable contribution of pathogenic microbiological material to streams and lakes both from point sources of treated agricultural wastes (e.g., meat processing plants, dairy sheds) and from diffuse sources of raw wastes from land-based animals (e.g., feral forest mammals, pastoral agriculture, birds). Sources of human wastes (e.g., from town oxidation ponds) make their own obvious contribution also. Such questions have been of concern for some time (e.g., for Lake Taupo, Miles 1963).

Elements of the agriculture industry were expressing interest in the possible effects of poor microbiological water quality on the production of farming operations, via impacts on the health of stock and, potentially, in terms of international trade barriers.

The Ministry of Health became involved in the programme, through its interest in recreational water quality (i.e., the health of swimmers) and drinking water quality.

The Ministry of Commerce also had an interest in this study, as did the Department of Conservation, from a tourism and recreation perspective. These two agencies supported the concept of the programme.

These elements were combined in a successful initiative from the Ministries for the Environment, Health and Agriculture,[1] and were supported by the Ministry of Commerce and the Department of Conservation. This initiative was for a "Freshwater Microbiological Research Programme", funded by the Ministry of Research Science and Technology, in 1997. Funding was obtained for a five-year period, from 1997 to 2001.

The aims of this programme is to undertake the necessary science to enable robust guidelines to be developed for:

- Bathing and contact recreation[2]
- Stock watering

A third objective is to improve the current Ministry of Health's guidelines for

- Drinking Water.

It is to be noted that these objectives refer only to *surface* waters (ponds, lakes, streams and rivers) and not to groundwaters. It was considered early on in the programme that there was little understanding of the microbiological quality of groundwater and that this topic was sufficiently complex that it warranted a research programme in its own right.

Also disease-causing or illness-causing agents are confined to *pathogenic* organisms (multiplying in the invaded host) and so do not include toxigenic organisms (such as cyanobacteria) or chemical agents. These agents can be of concern, but are beyond the scope of this programme.

[1] As of 1 March 1998, this is the Ministry of Agriculture and Forestry.

[2] The current version is at www.mfe.govt.nz/about/publications/water_quality/revised_guidelines.pdf; see also the supporting manual at www.mfe.govt.nz/issues/water/recguide_suppman.htm

The main study carried out under the programme has been a 15-month national survey, sampling every fortnight, of 10 microbiological health-risk indicators and pathogens at 25 freshwater recreational and drinking water abstraction sites. Twelve Regional Councils carried out the sampling, with four laboratories performing the tests. The results have formed the basis of the risk assessment presented herein. Associated studies have also been completed on stock-water drinking patterns and farm trough surveys, and some analysis of current Canadian studies on the effects of water quality on cattle.

1.2 Management of the Programme

A Management Group was set up to guide and oversee the design and implementation of the programme, using a Regional Council representative, government department staff, a consultant from the private sector, and two Crown Research Institutes (see Appendix A.1). It has commissioned a number of reports (e.g., on literature reviews, method development, design of appropriate studies, analysis of results).[3] A Programme Manual has also been kept and updated from time to time (McBride *et al.* 2000b). The Management Group met regularly through the course of the Programme.

National and overseas experts have reviewed proposed work.

- In March 1998, Mrs Janet Gough (Lincoln Environmental), provided a risk management discussion for the whole programme, recommending that it proceed as proposed, with regular review (Gough 1998).
- In December 1998, Mr E Pike (retired Chief Microbiologist at the Water Research Centre, Medmenham, UK) reviewed work at that date, consisting of the results of a preliminary microbiological study (McBride 1998) and the design of a full study, recommending that the full study proceed.

After the sampling programme had been completed and the results reported internationally (Till *et al*. 2000), a proposed risk analysis protocol was drafted and reviewed within the Management Group and also by two international experts (in April 2001):

- Dr Peter Teunis at the National Institute for Public Health and the Environment, Bilthoven, The Netherlands.
- Dr Paul Gale, Microbiologist, Water Research Centre, Medmenham, UK.

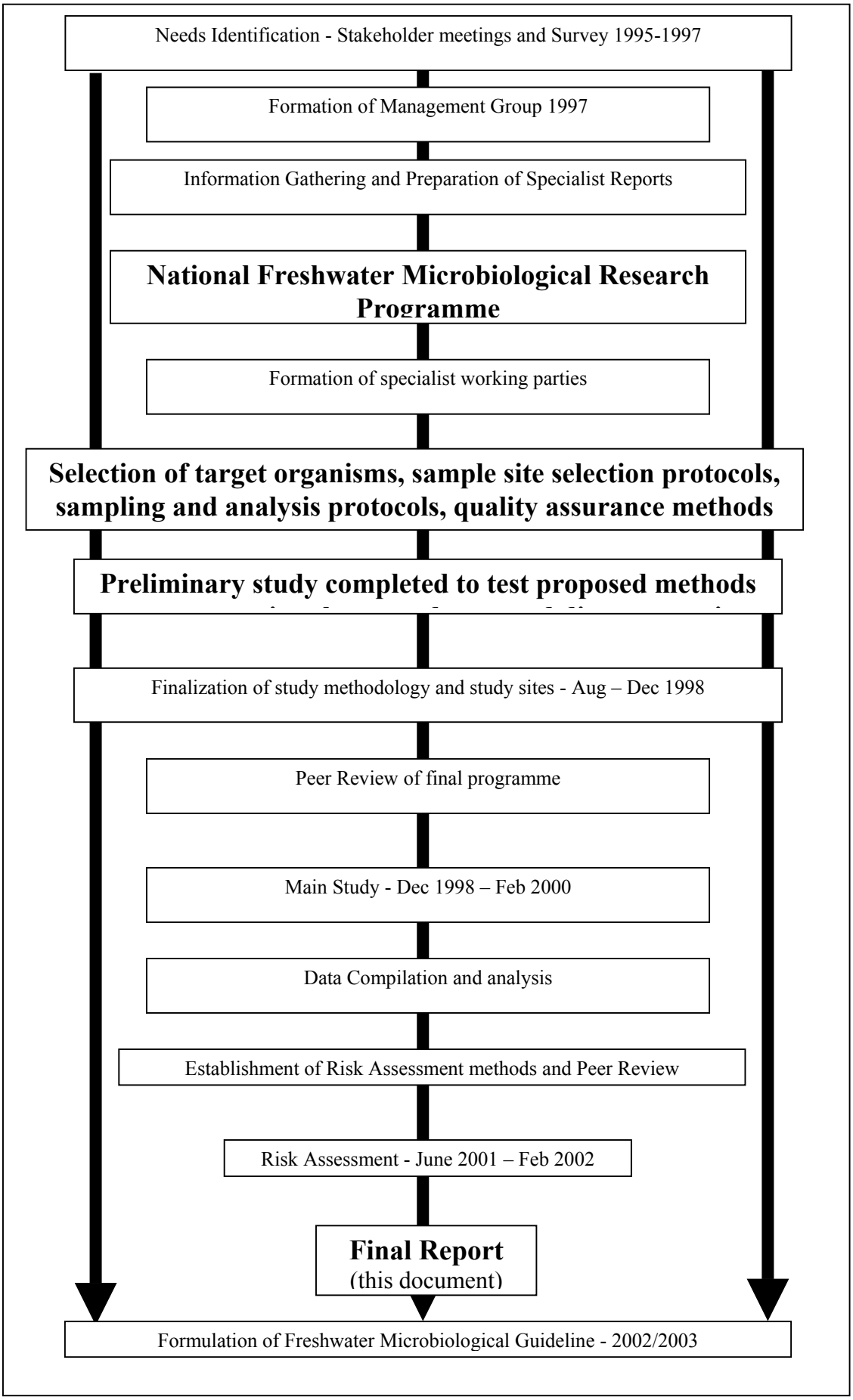
These reviews have resulted in the structure of this report.

Finally, the risk analysis work has been reported to and reviewed by a Risk Analysis Working Group (Appendix A.2).

A timeline for the Programme is given in Figure 1.

[3] Those commissioned by the Programme and referenced in this report are indicated by an asterisk in the references section, and are available on request from the Ministry for the Environment.

Figure 1: Overview of the Freshwater Microbiology programme



1.3 Overall Approach

No epidemiological studies have been undertaken. After consideration of ethical and logistical factors these were considered not feasible for the New Zealand freshwater environment (McBride *et al.* 1996).[4] Nevertheless the following model between indicators, pathogens and health risks has been adopted.

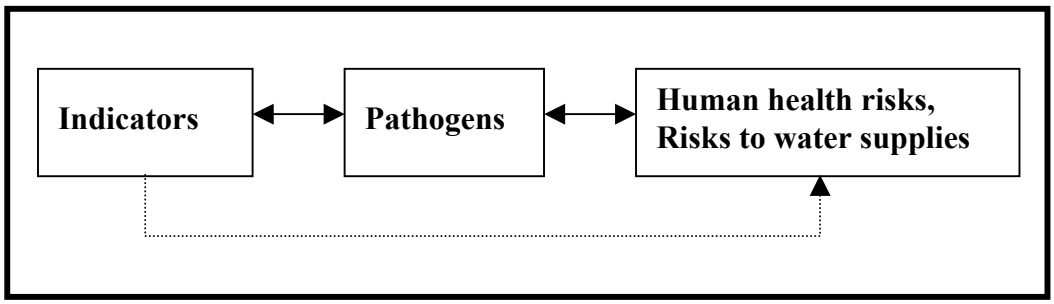


Figure 2. Overall programme model. Water managers want to understand the relationship between easily measured indicators and health effects (dotted line). The programme aims to establish the relationship between indicators, environments, and pathogens and use dose response models together with expert judgment to establish linkages between indicators and health effects.

Indicators may indicate the presence of particular pathogens, or their potential presence as a health risk. It is possible that they may be relatively poor indicators of a particular pathogen, yet still indicate a health risk (i.e., overall pathogenicity) generally. *E. coli* is used as the indicator of choice for example, as the indicator of faecal contamination of freshwater.

To achieve the aims of the programme we seek to establish relationships between:

a) Environmental sources of faecal pollution, levels of indicator microorganisms, and the potential for the presence of pathogens.

b) Environmental sources of faecal pollution, levels of indicator microorganisms, and the potential risk to human or animal health.

c) Environmental factors and land use that influence the presence of indicator microorganisms and pathogens in freshwater.

Some 25 sites in a range of catchment types were selected, the pathogens and indicators measured every two weeks over a 15-month period.

Where we have pathogen data they have been used to assess the relationships between possible health effects, pathogens and indicators (sufficient dose-response data have been found to perform quantitative risk assessments for infection). Where there are gaps in knowledge published information (e.g., Haas *et al*. 1999, Teunis & Havelaar 1999) has been used along with advice from the Risk Assessment Working Group. This includes advice on relationships

[4] A limited number of freshwater epidemiological studies have been undertaken overseas, e.g., in the Ardèche river basin in France (Ferley *et al*. 1989), North American lakes (Dufour 1984) and a rural USA stream (Calderon *et al*. 1991), as reviewed by Prüss (1998). In general these studies show an increasing illness risk to bathers as a function of the concentration of a bacterial indicator—especially *E. coli* faecal streptococci or enterococci. One should also note a recent German freshwater epidemiological study (Wiedenmann *et al*. 2002) in which *E. coli* has been linked to swimmers' illness risk (a full publication of this important study's results is yet to come). Marine studies have reached similar conclusions, e.g., in the USA, (Cabelli 1983, 1989; Haile *et al*. 1999), in the UK (Fleisher *et al.* 1996), and in New Zealand (McBride *et al*. 1998). Some freshwater studies have not reached such conclusions (e.g., lake studies by EHD 1980, Seyfried *et al*. 1985a&b, Lightfoot 1989, and the stream study by Calderon *et al*. 1991—although this last finding has been challenged, McBride (1993).

between the indicators[5] and pathogens in this study and related pathogens not included in the study. The expert working group has played a key role in this risk assessment.

Thus the programme's aim is to provide the scientific basis from the data collected to develop a decision support system that will enable water managers to estimate the pathogen levels at a particular site or stream. Water managers would use this information to assess:

- health risks to bathers.
- suitability for stock drinking water.
- the level of water treatment required for domestic and possibly industrial purposes.

1.4 Contents of this report

In this report we seek to analyse and report on available data and knowledge that bear on the objectives of this work. We do not propose explicit form for new guidelines for recreational water, drinking water and stock water. Rather, the information presented should form a helpful basis for the future formulation of such guidelines.

This report conveys the technical findings of the completed work, and some interpretation of it, all in the context of performing a risk analysis. Accordingly, in chapters 2 through 5 we outline the analysis framework being used, focussing on issues relating to human health. The manner in which these considerations gave rise to the work completed is presented as the framework is developed. In the interests of brevity, most technical details appealed to are contained in Appendices, and many footnotes have been used for minor detail. We have attempted to spell out major assumptions made in that analysis, as to the selection of data and methods for its analysis.

A discussion of options for risk management appears in chapter 6.

Stock drinking-water and potable drinking-water issues are presented in chapters 7 and 8, and the report finishes with Conclusions as Chapter 9.

[5] Indicators are not considered to be risks *per se*.

2 RISK ASSESSMENT FRAMEWORK

A number of options were considered for the risk assessment framework. In essence they all comprise a "… qualitative or quantitative characterization and estimation of potential health effects associated with exposure of individuals to hazards (materials or situations, physical, chemical and or microbiological agents" (Haas *et al*. 1999). The particular model followed here is that proposed for environmental issues following the recent model put forward in the UK explicitly for environmental risk assessment (DETR/EA/IEH 2000). This is a four-step process:

1. Hazard identification

2. Identification of consequences

3. Risk Characterisation—the programme aims (as in section 1.3) are included here.[6]

4. Risk management options appraisal.

We also draw on recent material from Standards Australia and Standards New Zealand.

The overall objective is to enable robust guidelines to be developed for bathing and contact recreation. Suitability for stock drinking water and water treatment requirements will be addressed in the risk assessment but are more appropriately considered as a risk management issue.

Full account is taken of current risk analysis literature as it relates to water in the environment.

[6] We define "risk" as the combination of the probability of occurrence of a hazard and the potential for infection (consequences would be illness); accordingly the phrase "Estimation of the probability of the risk occurring" (DETR/EA/IEH 2000) is inappropriate.

3 HAZARD IDENTIFICATION

With hazard defined as a "property or situation that could lead to harm" (DETR/EA/IEH 2000), or "source of risk", an encompassing term that includes all sources of risk where there is a cause-effect relationship (AS/NZS HB 203:2000), the hazard to be addressed is

> *The presence of pathogenic microorganisms of faecal origin from animals and humans in inland recreational waters (streams, rivers, pools and lakes*).

Having regard to the Programme's budget, and to the pattern of recreational freshwater usage (McBride *et al*. 1996), the Management Group decided on a microbiological survey at 25 freshwater recreational sites spread through the North and South Islands. The full programme design noted that that health records often indicate both seasonal and regional variations in waterborne illness rates (e.g., for campylobacteriosis, Hearnden *et al*., in prep.). Evidence to hand suggested substantial variability of microorganism concentrations from one week to the next, possibly in response to upstream rainfall-runoff patterns. Also, expert advice and some historical data was to the effect that different levels of contamination are to be expected in catchments according to their predominant land uses. Accordingly, the guiding notion was that of five main areas of the country and five catchment types, where the predominant uses affecting microbiological quality are: birds/wildfowl, dairying, forestry/undeveloped, municipal, and sheep/pastoral. Some sites have also been selected because drinking-water is abstracted from them and conveyed to a treatment plant before reticulation.[7] Fortnightly sampling over a 15 month period was selected to give coverage of two summer (recreational) seasons and to give some expectation of temporal variability.

Staff of 10 Regional Councils and 3 District Councils, as support-in-kind, competently carried out the sampling.

Many water-borne illnesses, candidate pathogens, microbiological methods and health risk indicators were considered (Ball 1997a&b, Ball *et al.* 1998a&b).[8] This included animal pathogens and possible effects on animal health and weight-gain rates (Belton 1997). An expert panel of microbiologists considered this review and selected indicators and pathogens of relevance to the study based on availability and robustness of methods and cost. In some cases methods had to be adapted for environmental samples and enumerative assays. These methods were subsequently validated in a preliminary survey May – August 1998.

[7] For this reason the hazard defined does *not* include possible pathogenic contamination of drinking water obtained from household taps.

[8] The potential use of faecal sterols was considered also (Ball 1998c), and samples have been stored for this purpose at some later date

4 IDENTIFICATION OF CONSEQUENCES

We define "consequence" as the outcome of an event/incident expressed qualitatively or quantitatively as a concern. There may be a range of possible outcomes associated with an event (AS/NZS 4360:1999, Risk Management). The "consequences" considered are

- *Infection and/or illness.*
- *Mildness of illness (for the average healthy adult-the group most estimates have been based on) as compared to severe health effects (immunocompromised/very young/very old populations) that could be experienced, including acute gastro-enteric symptoms vs. systemic illness.*
- *Infection of domestic animals using the water.*

and as a further consequence

- *The requirement to treat water intended for potable supplies.*

A fundamental issue to be resolved is whether the probability to be considered is that of infection or of illness—the infection risk for the general population is generally larger than the illness risk for that population.

The Risk Analysis Working Group has recommended that infection, rather than illness, be treated as the end-point of the analysis. This is particularly because there can be a substantial number of infected people not presenting symptoms, yet whose infected faecal wastes can be passed through the sewerage system and into receiving waters and then on to recreational water users who may become ill. This choice means we will not be able to utilise data from many epidemiological (e.g., Ferley *et al.* 1989) and outbreak studies (e.g., MacKenzie *et al.* 1994), as they are generally confined to measures of illness, not of infection. But we will be able to use dose-response data reported in controlled clinical trials as these typically assess both illness and infection. They are however restricted to healthy adults, e.g., for giardiasis (Rendtorff 1954a&b, Rendtorff & Holt 1954a&b); campylobacteriosis (Black *et al*. 1988); cryptosporidiosis (DuPont *et al.* 1995, Okhuysen *et al*. 1998, Messner *et al.* 2001).

The routes of infection or illness include both ingestion and inhalation (the latter is likely to be more apposite for water-skiers than for swimmers or bathers, and is documented as a route—Couch *et al*. 1966a, Baylor *et al.* 1977). Note that infectivity may be quite different for these routes.

The main consequence of exposure to the hazard is a number of people being struck with mild illness.[9] In a small minority of cases there is a possibility of severe health effects. There is clear evidence that this occurs with immune-compromised people (e.g., those already infected with the HIV virus.[10]) General understandings are also that babies and the very elderly may be at increased risk as would also be those on some medication/radiation treatment (i.e. cancer).

[9] For example, the following illness have been attributed to swimming in contaminated water: cryptosporidiosis (Sorvillo *et al*. 1992, Baker *et al*. 1998—this was in a Hutt Valley swimming pool), campylobacteriosis (Koenraad *et al*. 1997), Norwalk gastrointestinal illness (Barron *et al*. 1982), and Hepatitis A (Bryan *et al.* 1974). There is also clear New Zealand evidence of illness being caused by contaminated drinking water (campylobacteriosis, Briesman 1987, Stehr-Green *et al.* 1991; giardiasis, Fraser & Cooke 1991; cryptosporidiosis, Duncanson *et al.* 2000, Eberhart-Williams *et al*. 1997). Severe illness outbreaks among recreational water users can only be expected in the presence of grossly contaminated waters, e.g., typhoid in Alexandria, Egypt (El Sharkawi & Hassan 1979); paratyphoid in England and Wales (Public Health Laboratory Service 1959).

[10] In 1993 a substantial outbreak of cryptosporidiosis arose in Milwaukee (Wisconsin), caused by contaminated drinking water. Some 403,000 people were estimated to have developed symptoms (MacKenzie *et al.* 1994). Another analysis inferred that 50 deaths were caused, among immune-compromised folk (Hoxie *et al*. 1996). There is another (unsubstantiated) claim that more than 4,000 hospitalisations and 104 deaths were caused (Morris *et al*. 1996). It is to be noted that the fundamental paper on this outbreak (MacKenzie *et al.* 1994) does not contain any statements about mortalities, but is often cited as such! A later outbreak

Some seasonal variation of illness could occur. For example, influenza virus tends to occur predominately in autumn and winter, seasonal variation of campylobacteriosis changes from the north to the south of New Zealand (Hearnden *et al*., in prep.). Australian evidence[11] has it that enteroviruses are more common during the late summer and autumn.

The virus and protozoa data obtained during the FMRP surveys depend substantially on the method used. Explicit account needs to be taken of the recoveries obtained using these methods.

Finally, the issue of subsequent additional infections needs to be considered as a possible consequence; that is, persons contracting mild illness from freshwater recreation may then become more vulnerable to other (possibly more serious) illnesses derived from any source, or a pre-existing condition may become aggravated. This may be particularly the case with respiratory illnesses.

occurred in Las Vegas and was listed as a contributing cause of the deaths of at least 20 HIV-infected adults, via drinking water (Goldstein *et al*. 1996). See also reviews of outbreak data by Rose (1997) and Craun *et al.* (1998).

[11] Presented by Dr G.S. Grohmann (Environmental Pathogens Pty. Ltd., Sydney) to the Environment Court in 1997 in relation to appeals (subsequently withdrawn) against an ARC decision to grant consents to WaterCare for the present and future Mangere treatment plant effluent discharges to Manukau Harbour.

5 RISK CHARACTERISATION FOR HUMAN HEALTH

The three aims of this programme (Section 1.3), as they affect human health, are addressed in the following three sections, on: Potential for presence of pathogens; Potential for health effects; Environmental factors and land use.

5.1 Potential for presence of pathogens

This topic is approached by way of analysis of the surveys carried out over 1998-2000.

5.1.1 Preliminary survey

By mid-1998 the main strategy for a national survey had been mapped out, consistent with the budget and time available: i.e., monitoring about 10 indicators and pathogens at 25 sites fortnightly for 15 months. But first a set of preliminary surveys was mounted at three sites in May-August 1998 to validate field and laboratory testing methods to be used. The sites[12] were selected as representative of relatively un-impacted bathing waters, waters impacted by animal waste material, and waters impacted by human waste material. They were:

- Upper Ruamahanga River, at the Double Bridges bathing site, above Masterton. This is a short distance from the Tararua Forest Park; the upstream water is relatively free from major faecal input, but the catchment contains some sheep farming operations.
- Waimakariri (Canterbury), below the PPCS Freezing Works discharge and 0.4 km below the rail bridge.
- Waikato River, downstream of Hamilton City's Wastewater Treatment Plant's effluent discharge.

The main study design did not include sampling for short-term (within-day) variability. To check on the consequences of this the preliminary survey conducted sampling at two times (about 1100 and 1400 hours) on each of 10 days. Samples from upstream of the effluent discharges at the Waimakariri and Waikato River sites were also conducted on two occasions.

These surveys included methods for the following faecal indicators and pathogens:

Indicators: *Escherichia coli, Clostridium perfringens* spores, enteroviruses (also a pathogen), somatic coliphage, and FRNA phage.

Pathogens *Salmonella*, *Campylobacter*, *Giardia* cysts, *Cryptosporidium* oocysts, Human Adenoviruses, Human Enteroviruses.

The virus assays are not enumerative; they indicate presence or absence only.

Detailed results of this survey have been reported earlier (McBride 1998, MfE 1998a). Main findings are given in the following.

5.1.1.1 Method selection

The evaluation of the methods resulted in the following choices being made for the main study. Colilert™ Quantitray for *E. coli*; selective enrichment of *Campylobacter* using a 3x3x3 MPN format, followed by PCR to detect thermotolerant species; PCR for viruses following concentration using Virasorb filters; wound polypropylene thread cartridge filtration (CUNO MICROWYND) of a standard volume of 100 L and IMS (Immunomagnetic Separation) recovery for protozoan cysts. MPN tables were designed for the *Campylobacter and Salmonella* assays (McBride in prep.[13]). Detection limits are: *E. coli*, 1 /100 mL; *C. perfringens* spores, 1 /100 mL; FRNA and somatic phages, 1 /100 mL; *Campylobacter* 0.3 /100 mL; *Salmonella*, 1.2 /L; Human Adenoviruses and Human Enteroviruses, (detected/ not

[12] Maps and photographs of these sites are available on request.
[13] Following occupancy theory methods promoted by Tillett & Coleman (1985).

detected per litre, by PCR methods).[14] Details of these methods are available in the Manual of Microbiological Methods for the Freshwater Microbiology Programme (Donnison 1998a & b).

5.1.1.2 Results

While the primary aim of these preliminary surveys was to agree on a suite of sampling and laboratory methods over a range of potential contamination, a number of other findings did emerge, including:

- Differences between morning and afternoon sample results tended to be small.
- The Upper Ruamahanga site was relatively clean (as expected), with low levels of *Giardia* cysts, 6/20 positives for enteroviruses and no adenoviruses. *Campylobacter* was present in higher concentrations than anticipated.
- The Waimakariri site was more contaminated than the Upper Ruamanhanga: *Giardia* cyst (IMS) results were up to 50 per 100 litres. *Campylobacter* and *Salmonella* were often detected, with the former being in excess of 100 per litre on six of the 20 occasions. Enteroviruses were detected on 4/20 samples and adenoviruses on 13/20.
- The Waikato site was generally the most contaminated. *Giardia* cysts[15] were found in all samples, reaching 527 per 100 litres. *Campylobacter* were found in all samples, with two being in excess of 100 per litre. Enteroviruses were detected on 15/20 samples and adenovirus on 10/20.

5.1.2 Implementation of Full Survey

Following the preliminary study, 25 recreational and water supply sites were selected on the basis of representing five different categories of environmental impact (land use, waste discharge, and waterfowl activity).[16] Staffs of Regional Councils were widely consulted in selecting sites and in making this categorisation. These are **B** (birds), **D** (dairy farming), **F** (forestry/undeveloped), **M** (municipal), and **S** (sheep/pastoral). These categories reflect the dominant impact in the catchment, as reflected in the table 1 below. The site groupings are listed in Appendix A.3.2.

Table 1: Water sampling site categories for the Freshwater Microbiological study

Category	**Predominant Impact***	**OTHER POTENTIAL IMPACT ON SOURCE WATER**
Municipal (**M**)	Urban development	Influences from **D**, **F** and **S**
Birds (**B**)	Birds (seagulls), waterfowl	Influences from **M**, **F** and **S**
Dairying (**D**)	Dairy farming	Influences from **F** and **S**
Sheep/pastoral (**S**)	Sheep farming	Influences from **D** and **F**
Forestry/ Undeveloped (**F**)	Forestry, exotic and/or native	Influences from pastoral activities (small lifestyle blocks)

*Other potential influences impacting on source water were assessed as being within 2 kilometres upstream of a recreational site. As all the sampling points were selected as being at recognised recreational sites, toilet facilities were available at all the **D**, **S** and **F** sites, usually in the form of septic tank disposal.

Samples were analysed within 24 hours of collection by four nationally recognised laboratories. Three are "Crown Research Institute" laboratories: two within the Institute of Environmental Science and Research Ltd (at Christchurch and at Kenepuru—near Wellington), the other being at the Agresearch site at Horotiu—near Hamilton (previously

[14] Useful reviews of viruses in the New Zealand environment are presented by Lewis & Greening (2000) and by Murray *et al.* (2000).

[15] Enumerated using the IMS methodology.

[16] All but the Ashburton River site are at or very near to a recreational site. Five sites are at water supply abstraction points, including the Ashburton River. Maps and photographs of site locations are available on request.

operated by the Meat Industry Research Institute of N.Z.). The fourth laboratory (MicroAquaTech, Protozoa Research Unit, Massey University, Palmerston North) performed Protozoa analyses. To minimise inter-laboratory variability each indicator or pathogen was analysed in only one of these laboratories, except for *C. perfringens* and *E. coli*—samples from the South Island were analysed in a Christchurch laboratory and the North Island samples were analysed in Hamilton.

5.1.2.1 Microbiological methods

These have been detailed in full in the Manual of Microbiological Methods for the Freshwater Microbiology Programme (Donnison 1998a&b). They are essentially as described in section 5.1.1.1. It is to be noted that the primers used in the virus assays were human-specific, i.e., viruses of animal origin would not be discovered by these tests. Other pathogens (*Campylobacter*, *Salmonella*, *Giardia* cysts and *Cryptosporidium* oocysts) are found in animals, can infect humans, and are detected by the methods used. A summary of the microbiological analysis methods and laboratories used are shown in table 2.

Table 2: Microbiological analysis methods for Freshwater Microbiology study

Target Microorganism	Methods	Laboratory
E. coli	Colilert (100 mL sample)	ESR: Christchurch Science Centre. MIRINZ: Hamilton
Thermotolerant *Campylobacter*	2 Stage MPN (330 mL sample) in Preston's Broth detection by group specific PCR to identify *C. lari* or "other thermotolerant species". A subsequent PCR retest of positive samples detected *C. coli* and *C. jejuni*, but not *C. lari* or *C. upsaliensis.*	ESR: Christchurch Science Centre
Salmonella	MPN (830 mL sample) in peptone enrichment media, plate to RSV and Selenite Cystine broth, XLD agar and confirm by ELISA or Biochemical properties	ESR: Christchurch Science Centre
Clostridium perfringens spores	Membrane filtration on TCS-Fluorocult agar.	MIRINZ: Hamilton
Somatic coliphage	Single layer plaque assay. Host *E. coli* WG5	MIRINZ: Hamilton
FRNA Coliphage	Single layer plaque assay. Host *S .typhimurium* WG 49 (F'lac: TN5)	MIRINZ: Hamilton
Giardia cysts	Filtration, immunomagnetic separation of cysts and detection using fluorescent-labelled monoclonal antibody	MicroAquaTech: Massey University, Palmerston North
Cryptosporidium oocysts	Filtration, immunomagnetic separation of oocysts and detection using fluorescent-labelled monoclonal antibody	MicroAquaTech: Massey University, Palmerston North
Human Enterovirus Group	Virus capture on +ve charged filter, elution, concentration and nucleic acid extraction. Detection using virus group specific RT-PCR	ESR: Porirua Science Centre
Human Adenovirus Group	Virus capture on +ve charged filter, elution, concentration and nucleic acid extraction. Detection using virus group specific PCR	ESR: Porirua Science Centre

5.1.2.2 Sites and sampling

Staff of 12 Regional Councils, carried out sampling on 29 occasions, from December 1998 until February 2000 (with a one-month gap in July 1999). Only one sample per day was taken for the suite of analyses—the preliminary study had not revealed much evidence of with-day variability and so it was decided to maximise the spread of sites and sampling intensity within the available budget.

On each occasion samples had to be sent by courier to the 4 laboratories. This has resulted in a dataset of 725 values for each of the determinands. Diligence by all concerned has resulted in only one missing value in all the entire set of 7250 values (*Campylobacter* at the Lee River, on 3 May 1999[17]).

A detailed catchment assessment (sanitary survey) was conducted for each site, covering a minimum distance of 2 kilometres upstream for those sites on rivers. This was to check that the sites selected on information provided by Regional Councils were representative of the five different categories of environmental impact.

Mr D. Till maintained day-to-day supervision of the programme.

5.1.3 Main findings of the Full Survey

Details of the findings are given in Appendix A.3 (and in Till *et al*. 2002). In the following sections we discuss the broad features of the determinands measured. We then look at how they vary with respect to:

- Spatial patterns (between catchment types).
- Temporal patterns.
- Correlations determinands.
- Statistical modelling.

An exploratory data analysis approach is taken in the first three sections. It is only in the last category that we consider formal statistical hypothesis testing. We first examine the general pattern of results obtained from Appendices A.3.3–5, and then highlight any particular further features of temporal and spatial variations and correlations.

General patterns (Appendices A.3.3–5)

The 725 data collected for each variable data show that

- *E. coli* was nearly always detected (it was detected in 99% of all samples), with somatic coliphage being detected most of the time (89%). Both tended to be highest in summer/autumn.
- *Clostridium perfringens* spores and FRNA phage were detected in only about half of all samples (58% and 52% respectively). They tended to be elevated in winter.
- *Campylobacter* were detected in 60% of all samples. Six percent (i.e., 43) of all samples were above the test's detection limit (i.e., >110 per 100 mL). They were highest in late summer-early autumn. *C. jejuni* was the most frequent thermotolerant species identified, being present in at least 48% of the positive samples.
- Both virus groups were detected in about one third of all samples, though they were strongly dissociated (i.e., if adenovirus was detected in a sample enterovirus was seldom detected, and vice versa). No clear temporal pattern was obvious.
- The *Salmonella* detection rate was only 10%, and was highest in August.
- Cyst/oocyst detection rates were very low (8% and 5% respectively); cyst and oocyst detections were strongly dissociated.

[17] Some repeat sampling was necessary at the two Southland sites in September 1999 also.

- With some exceptions, correlations between indicators and pathogens were generally low. However, somatic coliphage was well correlated with *E. coli* and there were moderate correlations between somatic coliphage and *Campylobacter*, between somatic coliphage and FRNA phage, and also between *E. coli* and *Campylobacter*.[18]
- There was a lack of correlation between viruses and phages, between viruses and *C. perfringens* spores (as revealed by scatterplots and attempts at logistic modelling—not shown in the Appendix).
- Correlations between data collected within the bathing season are very similar to those obtained using all the data.

Spatial patterns (see Appendix A.3.3)

Appendix A.3.3 depicts the variation of bacteria, phages and cysts over the five catchment types, via boxplots. Virus results over the catchment types are shown as tables only—box plots cannot be calculated because these data are dichotomous (i.e., they are either present or absent). Similar tables are shown for the cysts to aid the interpretation of their boxplots, because the proportion of their positive results was so low.

The following general patterns of spatial variation over the five catchment types can be adduced (Table 3).

Table 3: Spatial patterns of occurrence of different groups of microbial pathogens and indicator organisms.

Organism	**%Detects**[+]	**Ranking**	**Notes**
E. coli	99	**B** > **D** ≈ **S** > **F** ≈ **M**	24 data greater than 2400/100 mL
C. perfringens spores	57	**B** ≈ **D** ≈ **S** ≈ **M** > **F**	Similar pattern to *E. coli*
Somatic coliphage	89	**B** ≈ **D** ≈ **S** > **F** ≈ **M**	Similar pattern to *E. coli*
FRNA phage	52	**B** ≈ **S** ≈ **D** > **M** > **F**	Few phages in **F** catchments
Salmonella	10	**S** > **B** > **D** ≈ **F** > **M**	%Detects: 21 (**S**), 14 (**B**), 7 (**D**), 2 (**F**), 0 (b)
Campylobacters	60	**S** ≈ **B** > **D** ≈ **F** > **M**	%Detects: 72 (**B**), 65 (**S**), 60 (**D**), 53 (**F**), 49 (**M**). But **S** catchments had the highest proportion of high values (i.e., ≥110 per 100 mL): **S** (13.6%) > **B** (9.5%) > **M** (9.2%) > **F** (5.9%) > **D** (4.1%). **M** showed the greatest variability[19]
Giardia cysts	8	**B** > **D** ≈ **F** ≈ **M** ≈ **S**	%Detects range: 15 (**B**) – 6 (**S**)
Cryptosporidium oocysts	5	**B** > **D** ≈ **F** > **S** ≈ **M**	%Detects range: 9 (**B**) – 2 (**S**)
Human Adenovirus	32	**B** > **S** > **F** > **M** > **D**	%Detects range: 44 (**B**) – 21 (**D**)
Human Enterovirus	33	**D** ≈ **B** > **M** ≈ S > **F**	%Detects range: 38 (**B**) – 29 (**F**)

[+]Percentage of positive results in the set of samples. Catchment types: **B** (birds), **D** (dairy farming), **F** (forestry/undeveloped), **M** (municipal), and **S** (sheep/pastoral). NB. Virus results between catchment types are considered similar if their proportion of detects differ by no more than 2%. Similarity/dissimilarity for the other microorganisms was judged by the closeness of their medians and similarity of their spreads.

[18] This inference is based on a calling a rank correlation coefficient between 0.4 and 0.7 "moderate", see the Spearman Rank Correlation table in section A.3.5.1.

[19] This is consistent with the sporadic occurrence of campylobacteriosis on communities—while there is a general overall presence of this disease, there are none-the-less occasional outbreaks (e.g., at swimming pools, Baker *et al.* 1998).

The main spatial features of these data are that:

- Bird catchments were the most contaminated, across nearly all microorganisms.
- Dairying catchments were often the second-most contaminated, but not for *Campylobacter* nor for adenoviruses.
- The Municipal (**M**) and Forestry/Undeveloped (**F**) catchments were generally the least contaminated.
- The general spatial pattern exhibited by *C. perfringens* spores and somatic coliphage was rather similar to *E. coli*, but with lower proportion of positive results.
- *Salmonella* is the only group for which sheep catchment types were the most contaminated.
- The overall pattern of distribution of *Campylobacter* species was similar between catchment types, except that the Sheep (**S**) catchments contained elevated levels of *C. lari* (33% of positive samples, versus 14% average for all other catchments).
- *Giardia* cysts and *Cryptosporidium* oocysts were detected rather infrequently and at rather low numbers in all catchment types, especially compared with the preliminary survey.[20]
- While both virus groups have similar overall proportion of positive results (about 33%), the range of adenoviruses detected among the catchment types (44%-21%) is larger than for the enteroviruses (38-29%).[21]

Temporal pattern (Appendix A.3.4)

Appendix A.3.4 depicts the temporal variation of bacteria, phages and cysts over the 15 months of the survey (with a one-month sampling gap in July 1999). These are presented as scatterplots of concentrations versus time. On each plot we have superposed a "smoother", the TREWESS smoother of Data Desk (Velleman 1997).[22] This technique allows us to discover trend patterns in data *without* having to specify in advance the functional form of that trend—classical techniques require us to specify this form, whereas a smoother "allows the data to speak for themselves".[23] The TREWESS smoother has been chosen because is appears to be the most versatile[24] in picking out time trends. The price one pays for this sensitivity is that a single very high result (e.g., FRNA phage in site type **B**) can cause a "blip", not necessarily reflecting an overall trend.

We can see the following time trends in the data:

E. coli	While tending to be to be highest in summer, there was no obvious pattern for **M** catchments
C. perfringens spores	Tended to be highest in winter.
Somatic coliphage	While generally highest in summer/autumn there was a tendency to be highest in **M** catchments in autumn and spring
FRNA phage	The winter peak for group **B** was due to a single sample. There is some evidence for high results in the first summer, but not in the second.
Salmonella	Highest in August, and in late summer for group **B**

[20] Some of this difference is attributable to the shifting of the preliminary survey's Waimakariri and Waikato sites upstream of point sources before their inclusion in the full study (so as to be at recreational sites). However, the preliminary study's Ruamahanga site was not shifted before its inclusion in the final study, and in the former it had 50% detects whilst in the full study it had only 4% (i.e., 2 detects in 27 samples). This suggests that there may have been generally less cysts present in the full survey than were present in the preliminary survey.

[21] A factor not considered at the time of site selection is that because all but two (Waiwhekaiho and Ashburton) are recreational sites (see Table A.3.1), they also have toilet facilities. Although facilities at Municipal sites are connected to city sewage systems, the catchment assessment identified the potential for an impact from combined stormwater/sewage overflows. Toilet facilities at other sites either used septic tanks or pit-privies for waste disposal that could account for the presence of human viruses.

[22] Other smoothers (e.g., LOWESS, moving averages,…) were tried but did not reveal the patterns shown by the TREWESS smoother.

[23] TREWESS = Trimmed Resistant Weighted Scatterplot Smooth. It accommodates unequally-spaced data and is thus suitable for smoothing scatterplots. It offers two parameters—the span of the smoother and the trimming percentage. We have used the default span of 20% of the data and the default trimming percentage is 10% trimmed mean. Setting the span larger makes the TREWESS smooth smoother and less willing to follow local fluctuations. Setting the trimming percentage larger makes TREWESS resistant to longer excursions in the data, but can also affect sensitivity and smoothness.

[24] Compared to other options, including LOWESS and MEDIAN smoothers.

Campylobacter	Highest in April in all groups. A distinct peak in the presence of *C. lari* occurred in July-August. Peaks in presence of *C. jejuni* were detected over the April-August period, and sometimes in summer.

Further examination of the data when split into three main regions (North, Central and South) did not reveal substantial between-region differences in pattern.

Correlations/associations among the determinands (Appendix A.3.5)

Appendix A.3.5 shows tables of correlations coefficients between determinands both over all sites, and within each catchment type. There are dangers in making the many comparisons that such tables invite, especially if one wishes to make statements about the "statistical significance" of each comparison (Scarsbrook *et al*. 2000). In this report we ignore such concerns.[25] Instead we seek the salient features of such tables. In doing so we recognise that this is an issue of substantial concern in the statistical literature, having a great deal to do with "burdens of proof". In effect we are taking a "face-value" stances on the interpretation of these data.

The Appendix also contains tables depicting the degree of agreement between all combinations of virus and cyst data.[26]

The main features of these data, beyond that already noted, are that:

- Correlations are influenced by catchment types.
- In the catchment groups with greater concentrations of *Campylobacter* (**S** and **M**) the correlation with *E. coli* is stronger, as evidenced by Spearman's correlation coefficients being above 0.5. (The largest proportion of high *Campylobacter* values occurred in the **S** catchments and the greatest spread of values occurred in the **M** catchments.)

General Linear Modelling (Appendix A.3.6)

We have used a combination of analysis of variance, analysis of covariance and logistic modelling, all of these are components of the "generalized linear model" (McCullach & Nelder 1995). The outcome of this modelling is a formal identification of the important variables explaining the patterns of microbiological contamination, as shown on Table 4.

We expected that turbidity would be better than river flow as an explicatory variable, because it picks out the rising limb of a flood where contamination is much higher than at same flow on falling limb (Nagels *et al.* in press). In fact the analysis found little difference between flow and turbidity as explicatory factors. On that basis, turbidity would appear to be the more practical variable, as it is by far the more easily measured or assessed.

[25] This is principally because the ability to pronounce "statistical significance" is predicated on the null hypothesis that in fact the true value of the coefficient is exactly zero, and then computing the probability of getting data at least as extreme as was obtained if that hypothesis were true. A "statistically significant" result is declared if that probability (the "*p*-value") is small enough (i.e., smaller than the *a priori* "significance level", typically taken as $\alpha = 0.05$). This declaration is made because it would appear to be unlikely to get such data *if* the tested hypothesis is true. But we do not hold this hypothesis to have ever been tenable in the first place! In particular, we disdain the practice, inherent in this approach, of declaring that there was "no correlation" if statistical significance has not been attained: one is never entitled to infer that such a null hypothesis is true (i.e., by "accepting" it, e.g., see Goodman 1993, 1999).

[26] In doing so, cyst data are reduced to a present/absent scale. The measure of agreement selected is known as "Cohen's kappa", as explained in that Appendix.

Table 4: Important environmental factors identified by GLM

Microorganism	***Important environmental factors***
E. coli	Turbidity (or flow), catchment type
C. perfringens spores	Turbidity (or flow)
Somatic phage	Turbidity (or flow), catchment type
FRNA	Turbidity (or flow), catchment type
Campylobacter	Turbidity (or flow), catchment type; or *E. coli*, within bathing season and within catchment type
Adenovirus	Catchment type
Enterovirus	Nothing
Cysts, *Salmonella*	Insufficient detects to make valid inferences

The result for *Campylobacter* is notable, in that if *E. coli* is included as a variable it turns out to have an important association with the levels of *Campylobacter*, and this effect differs strongly between catchment types and within or beyond the bathing season (tending to be higher in the bathing season). The converse is also true. This is at least partly explained by the pattern seen in the correlations; those between *Campylobacter* and *E. coli* are poor in catchment types where *Campylobacter* numbers tend to be low, but are higher when a greater spread or magnitude of them is found.

5.2 Potential for health effects

This section addresses the second programme aim: "Environmental sources of faecal pollution, levels of indicator microorganisms, and the potential incidence of human or animal health effects". The approach taken is given in details in Appendix A.3.7. In particular it is to be noted that for reasons given in Section 4 the analysis has focussed on infection rates, not on illness rates.

Note too that while six pathogens were assayed in the FMRP (*Campylobacter*, adenoviruses, enteroviruses, *Salmonella*, *Giardia* cysts and *Cryptosporidium* oocysts), the data collected demonstrated elevated levels of the first three only. We therefore confine the modelling to these three. Also, by making some assumptions about infection versus illness rates and concerning illness reporting rates we can make some calculations to translate the computed infection rate to an illness rate and compare the result with published figures for the national illness burden. This we can do only for campylobacteriosis because it is the only illness considered herein that is formally notifiable.

The calibrated dose-response parameters used to determine infection rates are based on clinical trials that have used selected pathogen strains prepared under prescribed methods, with in some cases limited numbers of subjects (e.g., for Adenovirus studies, Couch 1966a). In the few instances where more than one strain of a pathogen has been studied a wide range of infectivities has been reported (e.g., for cryptosporidiosis, Teunis *et al*. 2000a&b—see Table A3.7.1). It must therefore be recognised that a substantial degree of uncertainty may pertain to some of these estimated dose-response parameters.[27]

Essentially, because the focus of this analysis is on infection rates, we use existing data on dose-response modelling in a Monte-Carlo simulation of the infection rates, given inputs on:

- the duration of a swimming event

[27] A more formal investigation of these uncertainties is to be carried out, and will be reported in a further paper. This will include accounting for the precision of dose-response curves, following material in Teunis & Havelaar (2000).

- the volume of water ingested or inhaled per hour
- the microorganism concentration.

These inputs are "sampled" many times from appropriate probability distributions, using a Monte Carlo approach, to build up a distribution of results (rather than a single result).

The end result of this analysis is a set of risks for two target populations: a) individuals using a *particular* recreational site; and b) the population at large using a *multitude* of sites. These profiles, shown on Tables A3.7.3–5 are quite different. In the former case (a) the risk to an individual is usually very low, in that the median risk is zero.[28] This is because the water at the site is uncontaminated for a majority of the time. However on those few occasions where substantial contamination occurs, infection risks rise rapidly. In the latter case (b) there is always a minority of beaches that are contaminated and so there is *always* some risk of infection. The figures in the latter category can be used to estimate the proportion of notifiable illness cases that can be attributed to water contact recreation, as in the following section. We can also compute an average risk (i.e., spread over many exposures) for both cases.[29]

5.2.1 Campylobacteriosis

Of the three infections modelled (campylobacteriosis, adenovirus, enterovirus) only the first is a notifiable disease in New Zealand. The notified case rates for this illness are reported regularly in the New Zealand Public Health Reports, and have typically averaged about 300 per 100,000 per annum in recent years (approximately 400 in the summer months). From surveys of recreational water use (McBride *et al*. 1996) one can estimate that about 250,000 people go for at least one swim at a freshwater site each year (MfE 1998b). Further, most folk have been observed to immerse the head while swimming (McBride *et al.* 1996).

From the risk analysis reported herein, the median or mean campylobacteriosis infection rate spread over all recreational sites is approximately 0.04 (i.e., 40/1,000—see Table A3.7.3).[30] Therefore the typical number of infections per annum equates to 0.04 x 250,000 = 10,000. Accordingly, for the country's population of about 4 million, the water-recreation infection rate is 250 per 100,000 persons per annum.

If we assume that the notified illness rate reflects 13% of actual illness rate[31], the summertime illness rate is around 3,000 per 100,000 persons per annum. Furthermore, the infection rate is held to be double actual rate, i.e., 6,000 /100,000 persons per annum.[32]

Therefore the median proportion of Campylobateriosis illness that is attributable to freshwater contact recreation is 250/6000, i.e. 1 in 24, about 4%.[33]

[28] There is an argument in favour of using mean pathogen concentrations to calculate mean infection risk (Haas 1996), but in our context this applies to calculation of risks spread across beaches, not at an individual beach where risk percentiles are of direct interest.

[29] Using the approach advocated by Haas (1996), in which arithmetic average pathogen concentrations are used to calculate an average risk.

[30] This assumes that numbers of people actually swim on all occasions.

[31] As in a UK study (Wheeler *et al*. 1999), where the ratio of cases of campylobacteriosis in the community to that being notified was 7.6:1.

[32] As a generalisation, infection rate = symptomatic infection rate + asymptomatic infection rate. Asymptomatic infection rates are, for protozoa about 75% (Lopez *et al*. 1980), for bacteria between 50% and 66% (based on *Campylobacter jejuni* infection because it is the most notified—Figueroa *et al*. 1989, Ani *et al*. 1988), and viruses about 40% (Gianino *et al*. 2002) Because the disease burden is probably weighed with virus as first, bacteria second, and protozoa third in the number of cases caused, it seems reasonable to take an average figure of 50% asymptomatic infection rate for gastrointestinal pathogens generally. That is, the actual infection rate is double the symptomatic infection rate.

[33] If the ratio of notified:actual cases is as high as 50%, the attributable illness proportion would be 5 in 32 (i.e., 15%).

5.3 Environmental factors and land use

This section addresses the third programme aim: "Environmental factors and land use that influence the presence of indicator microorganisms and pathogens in freshwater". This will use the SPARROW suite of procedures (Alexander *et al*. in press, McBride *et al.* 2000a, Smith *et al*. 1997) to attempt to relate the pattern of microorganisms found with land use and catchment type.[34] This will attempt to determine the most important source of the microbiological data measured in the FMRP, and also to indicate what broad-scale changes to microbiological quality might occur if changes are made to these sources. [Note however that being a broad-scale (i.e., national scale) model it cannot be used to predict the consequence of local actions (e.g., retirement of stream riparian areas) as that must depend on local-scale studies.] This analysis will use land use data already obtained from AgriQuality's AgriBase dataset of the spatial distribution of land uses. Point source data has been obtained from Regional Councils. A digital elevation model has been developed from an enhancement of the contours data used to prepare the NZMS260 (1:50,000) map series—by "burning on" streams and lakes and forcing the contours into harmony with the flow of water.

Progress awaits the results of SPARROW modelling being carried out by NIWA and Rich Alexander (USGS, Reston, VA).

[34] The essence of the SPARROW technique (described in McBride *et al*. 2000 for an application to the Waikato catchment, see also Smith *et al.* 1997) is the fitting of a statistical model to monitoring data. In so doing this technique, being GIS-based and computer-intensive, takes explicit account of the dendritic drainage pattern of the landscape (through a digital terrain model), and so accounts for much more data than do standard statistical models (most of which do not incorporate an explicit spatial pattern). The fitting takes place by selecting many combinations of possible water quality determinants (land use, point sources, soil drainage) and seeing whether their inclusion in the model adequately improves its ability to match existing water quality data. Once a final model is thereby elected, it can be used to make predictions of future water quality.

6 RISK MANAGEMENT OPTIONS, RECREATIONAL WATERS

The end result of the risk characterisation is a prediction of health risk to a hypothetical group of individuals with stated assumed exposures (e.g., the quantity of water ingested or inhaled during water recreational activities). This can then be married with data on the actual recreational-use patterns and the levels of pathogens and indicators measured in the FMRP to arrive at a community health risk (and average risks spread over all individuals in a community).

"Risk Management" is the culture, processes and structures that are directed towards the effective management of potential opportunities and adverse effects. Several management options have been identified as the result of this study. These include:

- *Different intervention strategies to apply to various catchment area types that are likely to improve water quality (following SPARROW analysis).*

- *Changing current recreational water quality guidelines (presently being reviewed in light of the findings of the FRMP study).*

- *Alternative strategies to reduce the microbiological hazards to fresh waters (e.g. improved wastewater treatment and reducing the occurrence of stormwater into sewerage systems).*

- *Strategies to communicate risks to recreational water users.*

7 ISSUES FOR STOCK DRINKING WATER

The objectives of this Programme calls for consideration of the water resources used for stock drinking water. In doing so we have liaised with colleagues in Canada who have been carrying out detailed studies on the effects of different modes of drinking water supply on stock health.

7.1 Canadian studies

Canadian studies[35] on the water quality effects on cattle performance concluded that water quality could have a significant effect on livestock performance in that cattle having access to fresh water will consume more forage. It was quite obvious that cattle avoid water containing manure when they have a choice; and will refuse it when concentrations are too high at 0.025% fresh manure (Willms *et al*. 2000). Ongoing studies conducted over six years concluded additionally that cattle avoided water that was contaminated with 0.005% fresh manure by weight when given a choice of clean water and that clean water is essential for maximising cattle weight gains on Canadian rangeland during summer. Over a two-month period yearling cattle gained 23% more weight; and calves with dams 10% more weight when drinking clean water, than those drinking from a pond subject to stock exposure. The detrimental effect of pond water on weight gains of cattle appears to be mediated through feed intake rather than by stress induced by pathogens, toxins, or parasites. Cattle that drank clean water spent a longer time grazing and, in penned studies, ingested more food. Therefore, the proposed mechanism to explain weight gain response to water source appears to be defined by the palatability of water that influence water and forage consumption (Willms *et al*. 2001).

7.2 New Zealand studies

A New Zealand survey of stock drinking water on sheep, cattle and deer farms was undertaken with the primary objective of describing the water resources used for stock drinking water, and to elicit the opinions of farmers on the current quality of stock drinking water (Belton *et al*. 1998). This was in support of a national goal to establish stock drinking water quality guidelines.

The survey was stratified on a local government regional basis with a sample size of 85 farms per region; i.e., 1,190 properties nationally. The farm types were dairy, sheep, combined sheep/beef and deer.

Throughout New Zealand most farmers (approximately 90%) expressed satisfaction with quality of stock drinking water. About half considered that drinking water quality has an effect on animal production. Where there were problems, half were due to either low flows of water (resulting in stagnant water) or high flows (resulting in dirty water); a quarter were due to mineral contaminants and 20% were related to microbiological contaminants (either algae or effluent).

The survey data suggests three levels of utilisation of water resources for stock drinking water; (1) High use: Bore water (on 38% of properties this supplies 20% or more of water supplied). (2) Moderate use: Rivers, streams, springs dams and rural water schemes (10 to 25% of farms). (3) Limited use: Lakes, drains, water races, town water and rain water tanks (less than 5% of farms).

The survey concluded that farmers would support measures to protect the quality of water resources, but at the time of the survey (June 1998) considered microbiological contamination to be not a major concern. Current opinion could be different.

[35] The FMRP has had an involvement in the analysis of data from these studies (McBride *et al.* 1998).

A small farm trough water study investigated faecal contamination of trough water, by grazing livestock (Belton *et al*. 1999). Troughs are used for supplying drinking water to livestock on most New Zealand farms. The study confirms that faecal contamination of trough water supplies by grazing livestock is a normal occurrence on farms in which water is reticulated via troughs. Livestock contamination of trough water by regurgitation of rumen contents, and direct faecal splashing is likely to significantly affect trough water quality as measured by faecal coliforms, total coliforms and *Escherichia coli (E. coli*). The impact of this faecal contamination of trough water on animal production has not been measured in New Zealand.

Such contamination impacts on requirements or guidelines that may be set for management of sources of stock drinking water based on these indicator organisms -–if there is significant faecal contamination of water occurring in farm troughs, there may be little value in stipulating stringent water quality standards as measured by faecal coliforms, total coliforms or *E. coli.*

8 ISSUES FOR POTABLE DRINKING WATER

Of the 25 sites used in this study, five were source waters for treatment as community drinking-water supplies of which three were also recreational sites and all site types except forestry were represented. Viruses and *Campylobacter* were detected at least once at all sites. There was very little difference between the drinking-water supply sites and the remaining site types with respect to the occurrence of pathogens and the concentrations of indicator organisms.

The main issue for source waters is the high proportion of samples that contained *Campylobacter* (60%) and viruses (54%) and the ability of drinking-water treatment to kill or remove them. Drinking-water treatment tends to be optimised on the destruction of *E. coli* rather than pathogens. The median concentration of *E. coli* in the combined sites was 110/100mL.

A fundamental question is; do drinking-water treatment processes used in New Zealand ensure that, by the removal or destruction of *E. coli*, *Campylobacter* and viruses are also destroyed/removed.

9 DISCUSSION

A key outcome to be addressed in this Programme concerns the identification of the appropriate microbiological health-risk indicator to be used in future recreational water quality guidelines. In particular, should we change from the current choice (*E. coli*)[36] to something better? The results of this Programme do not suggest so. This is in spite of some obvious concerns.

First, *E. coli* have been shown to be capable of growth outside the gut under certain environmental conditions. *E. coli* have a temperature growth range between approximately 8 and 45°C with a doubling time ranging from 20 minutes to 25 hours (Bettelheim 1991) and may grow well where nutrient and water activity are adequate. *E. coli* growth has been reported in food (including *E. coli* O157:H7) (Doyle 1997), tropical water (Bermúdez and Hazen 1988), subtropical waters and soil (Hardina & Fujioka 1991), water in animal drinking troughs (Lejeune *et al*. 2001) and in temperate waters and sediments in water reservoirs near Sydney (N. Ashbolt, University of New South Wales, pers. comm.). It is possible that in warm weather some expansion of extraintestinal *E. coli* populations may occur in or near water bodies particularly where high organic carbon levels from vegetation or animal wastes are present. This may in part explain the observation that *E. coli* numbers in this study tended to be elevated in summer. If, under some conditions, *E. coli* proliferate in surface waters in New Zealand then its value as a faecal indicator is lowered as environmental growth of *E. coli* distorts the indicator/pathogen relationship. However, we have no evidence of this occurring during this study.

Second, studies have often found poor correlations between *E. coli* and particular pathogens (e.g., Borrego *et al.* 1987; Carter *et al*. 1987; Chauret & Armstrong 1995; Dutka *et al*. 1987; Sinton et al. 1993a&b[37]—see the review by Ball & Till 1998c). Yet, one may expect that *E. coli* may still serve as an indicator of health risk, rather than as an indicator of particular pathogens (i.e., the dotted line on Figure 2).

Other candidate indicators can be ruled out, as follows:

- FRNA phage was often not detected (as also reported for New Zealand marine waters by Lewis 1995), severely limiting its usefulness (it has often been advocated in overseas studies, e.g., Prof. M. Sobsey, University of North Carolina, Chapel Hill, pers. comm.)
- *C. perfingens* spores were also often not detected and tended to be highest in winter when *Campylobacter* was not peaking (increased runoff and sediment disturbance is a likely cause of this behaviour)
- Somatic coliphage was well-correlated with *E. coli* and its peak timed with *Campylobacter*, but its assay is more problematical.

A most useful result that has been found is a moderate correlation between *E. coli* and *Campylobacter*. That being so we can use the results of the infection risk analysis to make an assessment of a critical value for *E. coli*, corresponding to *Campylobacter* infection. This is based on the individual infection risk, i.e., the left-hand-side of Table A3.7.3. In that one sees that the breakthrough in infection risk occurs somewhere around the 75th percentile of the time. That is, infection risks are appreciable for the top quartile of the *Campylobacter* data. Accordingly, we can take that percentile of the *E. coli* data as a trigger point for *Campylobacter* infection—the appropriate percentile varies a little between catchment types, as shown by the shaded numbers in the Table A.3.3.2. From that Table trigger *E. coli* values are in the range of 175–500 per 100 mL. These values are not too dissimilar from results

[36] www.mfe.govt.nz/about/publications/water_quality/revised_guidelines.pdf

[37] Some of these studies have erroneously reported a finding of "no correlation"—because the data when subjected to a hypothesis test did not reject the hypothesis of "no association". This is a common logical error; all one can say in such circumstances is that the hypothesis was "not rejected". It is of course possible that the correlation may be low, or very low, but it is not tenable to claim it does not exist at all.

emerging from a recent German freshwater epidemiological study (Wiedenmann *et al.* 2002). They would need to be adjusted downward for the presence of other potential pathogens of course, and careful consideration would be needed regarding their statistical expression (e.g., as 95%iles, as is being advocated in the German study).

The risk analysis presented for campylobacteriosis must be treated with some caution—this is a developing field with authors as recent as 1999 noting that campylobacteriosis risks associated with untreated water being not yet quantified (Thomas *et al*. 1999). Also we note that the dose-response data used (Black *et al*. 1988) is for a particular strain and that the lowest dose is not much less than the ID_{50} calculated by fitting a (beta-Poisson) model to those data (as noted by Savill *et al*. 2000, 2001).[38] Yet it is the best approach available at present. And given the fact that another New Zealand study has reported finding *Campylobacter* in 60% (i.e., 18/30) of surface water samples (Savill *et al*. 2001), this is an issue that cannot be treated lightly. At some stage issues of "acceptable" or "tolerable" risks will need to use such an analysis as a guide.[39] And the assumptions leading to the suggestion made (in Section 5.2.1) that about 4% of campylobacteriosis cases could be attributable to water contact recreation need careful evaluation.

This analogy between *E. coli* and pathogens cannot be carried out for the other significant pathogen identified in this study (Human adenovirus). The risk analysis carried out for adenovirus does however suggest that it poses risks similar to that posed by *Campylobacter*. Human enterovirus appears to be rather less infective than adenovirus, so while they were detected in similar proportions it appears that adenovirus is of more concern. The strong dislocation found between these virus groups has been found elsewhere (Prof. M. Sobsey, University of North Carolina, Chapel, Hill, pers. comm.).

The low detection rates encountered for *Giardia* cysts and *Cryptosporidium* oocysts is a surprise, given historical New Zealand data (Brown *et al*. 1992; Ionas 1997). This raises the prospect of cyclical patterns of their concentration over annual periods, a topic that seems worthy of further attention given the world-wide concern about them (in particular, *Cryptosporidium*).[40]

Salmonella was detected rarely in this survey. Of the 69 samples in which salmonellae were isolated, 49 were serotyped, including *S. brandenburg* (21), *S. typhimurium* (10), *S. hindmarsh* (3), *S. mississippi* (1) and *Salmonella* rough (1) with 20 unable to be typed. *Salmonella* tended to cluster in particular sites, specifically those in the southern region. This occurrence was most likely associated with outbreaks of *Salmonella brandenburg* reported in sheep and also cases in the human population in the southern region at that time.

[38] ID_{50} is the dose at which 50% of an exposed population becomes infected. This quantity is sometimes called the median infectious dose (e.g., Haas *et al.* 1996), which can be slightly confusing—it is not the median of the dose, but the dose at which 50% become infected.

[39] Mara (2000) points out the strong differences between the UK and USA on this matter.

[40] We note that very low levels were also obtained by Simmons *et al.* (2001) in a New Zealand study of roof-collected rainwater in the period 1996-1998, although other microbiological contamination was present.

10 CONCLUSIONS AND RECOMMENDATIONS

The present New Zealand Recreational Freshwater Guidelines are based primarily on trigger levels of *E. coli*. This study has demonstrated that *E. coli* concentrations alone are not sufficient to enable the health risk from recreational use of fresh waters to be assessed. This study has identified a number of factors that may be useful indicators of health risk (e.g., turbidity, catchment type etc.) that warrant further exploration.

- The present New Zealand Recreational Freshwater Guidelines need to be reviewed.

Drinking-water treatment tends to be optimised on the destruction of *E. coli* rather than pathogens. As a high proportion of samples in this survey contained *Campylobacter* (60%) and viruses (54%), the question is raised as to whether drinking-water treatment processes used in New Zealand ensure that, by the removal or destruction of *E. coli*, *Campylobacter* and viruses are also destroyed/removed.

- Drinking-water treatment processes used in New Zealand should be reviewed to ensure that they are capable of protecting consumers from waterborne campylobacteriosis and viral infections.

Technical limitations precluded the analysis in this study for some important waterborne pathogens (e.g. Norwalk-like viruses).

- Consideration be given to test stored virus concentrates retained during this study for the presence of Norwalk-like viruses.

The influence of a catchment on water quality cannot be fully assessed until completion of the SPARROW studies.

- Development of management guidelines, including options for remedial actions on catchments, to be assessed following completion of the SPARROW studies.

The survey of water quality and resources used for stock drinking water concluded that farmers would support measures to protect the quality of water resources, but at the time of the survey (June 1998) considered microbiological contamination to be not a major concern. This study has demonstrated (SPARROW findings to follow) that the distribution of pathogens via waterways is an issue.

- That an information package of these findings be developed and communicated to farmers.

11 ACKNOWLEDGEMENTS

As well as the members of the Management Committee and RAWG (see Appendices 1 & 2), we thank:

Samplers

Diane Orr	Northland Regional Council
Paul Smith	Environment Waikato
Ian Buchanan	Environment Waikato
Dr Joanne Deely	Environment Bay of Plenty
Brett Osborne	Environment Bay of Plenty
Tony Hickling	Environment Bay of Plenty
Melissa Penn	Taranaki Regional Council
Graham Sevicke-Jones	formerly Hawkes Bay Regional Council, now Wellington Regional Council
Anita Anderson	Hawkes Bay Regional Council
Sarah Pond	Horizons.mw
Brenda Rosser	Wellington Regional Council
Brett Stansfield	formerly Wellington Regional Council (Wairarapa Division), now Hawke's Bay Regional Council
Jane MacGibbon	Tasman District Council
Melanie Eccles	Marlborough District Council
Lynda Neame	Marlborough District Council
Julie Edwards	Canterbury Regional Council
Shirley Hayward	Canterbury Regional Council
Richard Durie	Ashburton District Council
Ali (Alastair) MacArthur	Otago Regional Council
Pete Stevenson	Otago Regional Council
Chris Jenkins	Southland Regional Council

Laboratory staff

Antony Pita	MicroAquaTech, Palmerston North
Rebecca Pattison (*nee* Cowie)	MicroAquaTech, Palmerston North
Andrea Donnison	Agresearch, Hamilton
Robyn Clemens	Agresearch, Hamilton
Gail Greening	ESR, Keneperu Science Centre, Porirua
David Hart	ESR, Keneperu Science Centre, Porirua
Marilyn Grubner	ESR, Keneperu Science Centre, Porirua
Anita Davies	ESR, Christchurch Science Centre
Marion Savill	ESR, Christchurch Science Centre
Chris Graham	ESR, Christchurch Science Centre

Reviewers

Dr Edmund Pike	formerly Water Research Centre, Medmenham, England
Dr Paul Gale	Water Research Centre, Medmenham, England
Dr Peter Teunis	National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands
Dr Janet Gough	Environmental Risk Management Authority, Wellington (formerly Lincoln Environmental, Christchurch)

12 REFERENCES

An asterisk (*) beside the first authors' name denotes a study commissioned by the FMRP.

Akin, E.W. 1981. Presented at the U.S. EPA Symposium on Microbial Health Considerations of Soil Disposal of Domestic Wastewaters (cited by Haas *et al*. 1999).

Alexander, R.B.; Elliott, S.; Ude Shakar; McBride, G.B. in press. Estimating the sources of nutrients in the Waikato River Basin, New Zealand. *Water Resources Research.*

Ani, E.A.; Takahashi, T.; Shonekan, R.A. 1988. *Campylobacter jejuni* antibodies in Nigerian children. *Journal of Clinical Microbiology 26(3)*: 605–606.

AS/NZS HB 203:2000. *Environmental Risk Management—Principles and Processes*. Standards Australia and Standards New Zealand document.

AS/NZS 4360:1999. Risk Management. Standards Australia and Standards New Zealand Standard.

Bacon, D.; Ewing, C.P.; Burr, D.H.; Rollins, D.M.; Guerry, P. 1999. Differences in virulence between *Campylobacter jejuni* 81-176 and NCTC 11168. 10th International Workshop on CHRO (*Campylobacter*, *Helicobacter*, and Related Organisms), 12–16 September 1999, Baltimore MD, USA. Notes p. 95.

*Ball, A. 1997a. Pathogens in fresh waters: preliminary scoping study. Client Report **CSC97/3**. Institute of Environmental Science and Research Ltd. (ESR). Prepared for the Ministry for the Environment. March. 21 p.

*Ball, A. 1997b. Options for microbiological methods.. Client Report **CSC97/20**. Institute of Environmental Science and Research Ltd. (ESR). Prepared for the Ministry for the Environment. March. not paginated.

*Ball, A.; Till, D.G. 1998a. Review of potential waterborne pathogens in New Zealand. Client Report **CSC97/22**. Institute of Environmental Science and Research Ltd. (ESR). Prepared for the Ministry for the Environment. April. 8 p.

*Ball, A.; Till, D.G. 1998b. Review of the relationship between pathogens and indicators in fresh water. Institute of Environmental Science and Research Ltd. (ESR). Prepared for the Ministry for the Environment. June. 13 p.

*Ball, A.; Till, D.G. 1998c. Faecal sterols as faecal source indicators. Client Report **CSC98/04**. Institute of Environmental Science and Research Ltd. (ESR). Prepared for the Ministry for the Environment. April. 6 p.

Baker, M.; Russell, N.; Roseveare, C.; O'Hallahan, J.; Palmer, S.; Bichan, A. 1998. Outbreak of cryptosporidiosis linked to Hutt Valley swimming pool. *The New Zealand Public Health Report 5(6)*: 41–45.

Barron, R.C.; Murphy, F.D.; Greenberg, H.B.; Davis, C.E.; Bregman, D.J.; Gary, G.W.; Hughes, J.M.; Schonberger, L.B. 1982. Norwalk gastrointestinal illness. An outbreak associated with swimming in a recreational lake and secondary person-to-person transmission. *American Journal of Epidemiology 115(2)*: 163–172.

Baylor, E.R.; Baylor, M.B.; Blanchard, D.C.; Syzdek, L.D.; Appel, C. 1977. Virus transfer from surf to wind. *Science 198*: 575–580.

*Belton, D. 1997. Animal pathogens report. Institute of Environmental Science and Research Ltd. (ESR). Prepared for the Ministry for the Environment. December. 8 p.

*Belton, D.; Ryan, T.; Irwin, G.; Cameron, C.; Duganzich, D. 1998. "National stock drinking water telephone survey June 1998." MAF Quality Management (Ministry of Agriculture and Forestry), Ruakura Research Centre, Hamilton. Prepared for the Ministry for the Environment. June. 28 p.

*Belton, D.; Irwin, G.; Donnison, A. 1999. Microbiological Freshwater Programme. Farm Trough Water Study. .Prepared for the Ministry for the Environment.

Berger, J.O.; Sellke, T. 1987. Testing a point null hypothesis: the irreconcilability of *P* values and evidence. *Journal of the American Statistical Association 82*: 112–139.

Bermúdez, M.; Hazen, T.C. 1988. Phenotypic and genotypic comparisons of *Escherichia coli* in pristine waters. *Applied and Environmental Microbiology 54(4)*: 979–983.

Bettelheim, K.A. 1991. The Genus Escherichia. *in* "The Prokaryotes" 2nd edition (eds A Balows *et al*.), Chapter 142, pp 2696–2736.

Bishop, Y.M.M.; Fienberg, S.E.; Holland, P.W. 1975. *Discrete Multivariate Analysis*. The MIT Press, Cambridge, Mass.

Black, R.E.; Levine, M.M.; Clements, M.L.; Highes, T.P.; Blaser, M.J. 1988. Experimental *Campylobacter jejuni* infection in humans. *Journal of Infectious Diseases 157*: 472–479.

Borrego, J.J.; Moriñgo, M.A.; de Vicente, A.; Córnax, R.; Romero, P. 1987. Coliphages as an indicator of faecal pollution in water, its relationship with indicator and pathogenic microorganisms. *Water Research 21(12)*: 1473–1480.

Brieseman, M.A. 1987: Town water supply as the cause of an outbreak of *Campylobacter* infection. *New Zealand Medical Journal 100*: 212–213.

Brown, T.J.; Hastie, J.C.; Kelly, P.J.; van Duivenboden, R.; Ainge, J.; Jones, N.; Walker, N.K.; Till, D.G.; Sillars, H.; Lemmon, F. 1992. Presence and distribution of *Giardia* cysts in New Zealand waters. *New Zealand Journal of Marine and Freshwater Research 26*: 279–282.

Bryan, J.A.; Lehmann, J.D.; Setiady, I.F.; Hatch, M.H. 1974. An outbreak of Hepatitis-A associated with recreational lake water. *American Journal of Epidemiology 99(2)*: 145–154.

Cabelli, V.J. 1983. Health effects criteria for marine recreational waters. Report EPA 600/1-80-031, USEPA Cincinnati, OH (first published 1980). NTIS access #: PB 83-259994).

Cabelli, V.J. 1989. Swimming-associated illness and recreational water quality criteria. *Water Science & Technology 21(2)*: 13–21.

Calderon, R.L.; Mood, E.W.; Dufour, A.P. 1991. Health effects of swimmers and nonpoint sources of contaminated water. *International Journal of Environmental Health Research 1*: 21–31.

Carter, A.M.; Pacha, R.E.; Clark, G.W.; Williams, E. A. 1987. Seasonal occurrence of *Campylobacter* spp. in surface waters and their correlation with standard indicator bacteria. *Applied Environmental Microbiology 53(3)*: 523–526.

Chappell, C.L.; Okhuysen, P.C.; Sterling, C.R.; DuPont, H.L. 1996. *Cryptosporidium parvum:* intensity of infection and oocyst excreta patterns in healthy volunteers. *The Journal of Infectious Diseases 173*: 232–236.

Chauret, C.; Armstrong, N.; Fisher, J.; Sharma, R.; Springthorpe, S.; Sattar, S. 1995. Correlating Cryptosporidium and Giardia with microbial indicators. *Journal of the American Water Works Association 87(11)*:76–84.

Cohen, J. 1960: A coefficient of agreement for nominal scales. *Educational and Psychological Measurement 20(1)*: 37–46.

Couch, R.B.; Cate, T.R.; Gerone, P.J.; Fleet, W.; Lang, D.; Griffithm W.; Knight, V. 1965. Production of illness with a small-particle aerosol of Coxsackie A_{21}. *Journal of Clinical Investigation 44(4)*: 535–542.

Couch, R.B.; Cate, T.; Douglas, R.G.Jnr.; Gerone, P.J.; Knight, V. 1966a. Effect of route of inoculation on experimental respiratory viral disease in volunteers and evidence for airborne transmission. *Bacteriological Reviews 30(3)*: 517–531 (includes discussion).

Couch, R.B.; Cate, T.; Fleet, W.F.; Gerone, P.J.; Knight, V. 1966b. Aerosol-induced adenoviral illness resembling the naturally occurring illness in military recruits. *American Review of Respiratory Diseases 93(4)*: 529–535.

Couch, R.B.; Knight, V.; Douglas, R.G.Jnr.; Black, S.H.; Hamory, B.H. 1969. The minimal infectious dose of adenovirus Type 4; the case for natural transmission by viral aerosol. *Transactions of the American Clinical Climatological Association 80*: 205–211.

Crabtree, K.D.; Gerba, C.P.; Rose, J.B.; Haas, C.N. 1997. Waterborne adenovirus: a risk assessment. *Water Science & Technology 35(11–12)*: 1–6.

Craun, G.F.; Hubbs, S.A.; Frost, F.; Calderon, R.L.; Via, S.H. 1998. Waterborne outbreaks of cryptosporidiosis. *Journal of the American Waterworks Association 90(9)*: 81–91.

Department of Health. 1992. Provisional microbiological water quality guidelines for recreational and shellfish-gathering waters in New Zealand. Department of Health, Public Health Services, Wellington, January.

DETR/EA/IEH 2000. Guidelines for Environmental Risk Assessment and Management Department of the Environment, Transport and the Regions, Environment Agency, and Institute for Environment and Health, UK.

DRG 2001. Pilot Study: Pilot Plant Investigations. Surrogate Study Results and Recommendations. Disinfection Review Group report to Watercare Services Ltd., Auckland. Final Draft for Review.

*Donnison, A. 1998a. Manual of microbiological methods: A list of methods used by the microbial analysis group for The Fresh Water Microbiology Programme. Meat Industry Research Institute of New Zealand Ltd. (Food Technology & Research). Prepared for the Ministry for the Environment. June. 8 chapters plus Appendix.

*Donnison, A. 1998b. Manual of microbiological methods for The Fresh Water Microbiology Programme. MIRINZ Food Technology & Research Ltd. Prepared for the Ministries of Environment, Health & Agriculture.August. 54 p.

Doyle, M.P.; Zhao, T.; Meng, J.; Zhao, S. 1997: *Escherichia coli* O157:H7. in Food Microbiology: Fundamentals and Frontiers. ASM Press. (ed. M.P. Doyle), pp. 171–191.

Dufour, A.P. 1984. Health effects criteria for fresh recreational waters. *EPA-600/1-84-004.* Cincinnati OH: US Environmental Protection Agency.

Duncanson, M.; Russell, N.; Weinstein, P.; Baker, M.; Skelly, C.; Hearnden, M.; Woodward, A. 2000. Rates of notified cryptosporidiosis and quality of drinking water supplies in Aoteroa, New Zealand. *Water Research 14(15)*: 3804–3812.

Dupont, H.L.; Chappell, C.L.; Sterling, C.R.; Okhuysen, P.C.; Rose, J.B.; Jakubowski, W. 1995. The infectivity of *Cryptosporidium parvum* in healthy volunteers. *New England Journal of Medicine 332(13)*: 855–859.

Dutka, B.J.; El Shaarawi, A.; Martins, M.T.; Sanchez, P.S. 1987. North and South American studies on the potential of coliphage as a water quality indicator. *Water Research 21(9)*: 1127–1134.

Eberhart-Phillips, J.; Walker, N.; Garrett, N.; Bell, D.; Sinclair, D.; Rainger, W.; Bates, M. 1997. Campylobacteriosis in New Zealand: results of a case-control study. *Journal of Epidemiology and Community Health 51*: 686–691.

EHD 1980. A study of disease incidence and recreational water quality in the Great Lakes Phase I. Report 81-EHD-67. Environmental Health Directorate, Health Protection Branch, Health and Welfare, Ottawa, Canada (prepared by The University of Toronto).

El-Sharaawi, A.H. 1985. Some goodness-of-fit methods for the Poisson plus added zeros distribution. *Journal of Applied Microbiology 49(5)*: 1304–1306.

El-Sharkawi, F.; Hassan, M.N.E.R. 1979. The relation between the state of pollution in Alexandria swimming beaches and the occurrence of typhoid among bathers. *Bulletin of the High Institute of Public Health of Alexandria IX*: 337–351 (reprinted by the Alexandria University Press, 1980).

Esrey, S.A.; Feachem, R.G.; Hughes, J.M. 1985. Interventions for the control of diarrhoeal diseases among young children: improving water supplies and excreta disposal facilities. *Bulletin of the World Health Organisation 63(4)*: 757–772.

Ferley, J. P.; 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. *International Journal of Epidemiology 18*: 198–205.

Fewtrell, L.; Bartram, J. 2001. *Water Quality: Guidelines, Standards and Health*. IWA Publishing Cornwall, UK.

Figueroa, G.; Galeno, H.; Troncoso, M.; Toledo, S.; Soto, V. 1989. Prospective study of *Campylobacter jejuni* infection in Chilean infants evaluated by culture and serology. *Journal of Clinical Microbiology. 27(5)*: 1040–1044.

Fleisher, J.M., Kay, D., Salmon, R.L., Jones, F., Wyer, M.D. and Godfree, A.F. 1996. Marine waters contaminated with domestic sewage: nonenteric illnesses associated with bather exposure in the United Kingdom. *American Journal of Public Health 86(9)*: 1228–1234.

Fleiss, J. L. 1981: *Statistical methods for rates and proportions*. 2nd ed. Wiley: New York.

Fraser, G.; Cooke, K.R. 1991. Endemic giardiasis and municipal water supply. *American Journal of Public Health 81(6)*: 760–762.

Gale, P. 1996. Developments in microbiological risk assessment models for drinking water—a short review. *Journal of Applied Bacteriology 81*: 403–410.

Gale, P.; van Dijk, P.A.H.; Stanfield, G. 1997. Drinking water treatment increases microorganism clustering; the implications for microbiological risk assessment. *Journal of Water Supply: Research and Technology – Aqua 46(3)*: 117–126.

Gale, P. 1999. Assessing the risk of cryptosporidiosis. *Journal of the American Water Works Association 91(3)*: 4–5.

Gale, P. 1998. Simulating *Cryptosporidium* exposures in drinking water during an outbreak. *Water Science & Technology 38(12)*: 7–13.

Gale, P.; Stanfield, G. 2000. Cryptosporidium during a simulated outbreak. *Journal AWWA, 92(9)*: 105–116.

Gerba, C.P.; Rose, J.B.; Haas, C.N.; Crabtree, K.D. 1996. Waterborne rotavirus: a risk assessment. *Water Research 30(12)*: 2929–2940.

Gianino, P.; Mastretta, E.; Longo, P.; Laccisaglia, A.; Sartore, M.; Russo, R.; Mazzaccara, A. 2002. Incidence of nosocomial rotavirus infections, symptomatic and asymptomatic, in breast-fed and non-breast-fed infants. *Journal of Hospital Infection. 50(1)*:13–17.

Gilbert, R.O. 1987. *Statistical methods for environmental pollution monitoring*. New York: Van Nostrand Reinhold.

Goldstein, S.T.; Juranek, D.D.; Ravenbolt, O.; Hightower, A.W.; Martin, D.G.; Mesnik, J.L.; Griffiths, S.D.; Bryant, A.J.; Reich, R.R.; Herwaldt, M.D. 1996. Cryptosporidiosis: an outbreak associated with drinking water despite state-of-the-art water treatment. *Annals of Internal Medicine 124(5)*: 459–468.

Goodman, S.N. 1993. *p* values, hypothesis tests, and likelihood: implications for epidemiology of a neglected historical debate. *American Journal of Epidemiology 137(5)*¨495–501.

Goodman, S.N. 1999. Toward evidence-based medical statistics. 1: The *P* values fallacy. *Annals of Internal Medicine 130(12)*: 995–1004.

*Gough, J. 1998. Risk review of the freshwater research microbiological programme. Lincoln Environmental. Prepared for the Ministry for the Environment. February. 9 p.

Graham, D.Y.; Jiang, X.; Tanaka, T.; Opekun, A.R.; Madore, H.P.; Estes, M.K. 1994. Norwalk virus infection of volunteers: new insights based on improved assays. *Journal of Infectious Diseases 170*: 34–43.

Haas, C.N 1983. Estimation of risk due to low doses of microorganisms: a comparison of alternative methodologies. *American Journal of Epidemiology 118(4)*: 573–582.

Haas, C.N. 1996. How to average microbial densities to characterize risk. *Water Research 30(4)*: 1036–1038.

Haas, C.N. 1999. On Modeling Correlated Random Variables in Risk Assessment. *Risk Analysis 19(6)*:1205–1214.

Haas, C.N.; Crockett, C.S.; Rose, J.B.; Gerba, C.P.; Fazil, A.M. 1996. Assessing the risk posed by oocysts in drinking water. *Journal of the American Water Works Association 88(9)*: 131–136.

Haas, C.N.; Rose, J.B.; Gerba, C.P. 1999: *Quantitative Microbial Risk Assessment*. John Wiley, New York, 449 p.

Haile, R.W.; Witte, J.S.; Gold, M.; Cressey, R.; McGee, C.; Millikan, R.C.; 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(4)*: 355–363.

Hardina C.M.; Fujioka, R.S. 1991. Soil the environmental source of *E.coli* and enterococci in Hawaii's streams. *Environmental Ecotoxicology and Water Quality 6*: 185–195.

Hearnden, M.; Skelly, C.; Eyles, R.; Weinstein, P. in prep. Space-time dimensions of campylobacteriosis in New Zealand: A new window on disease ecology.

Hornick, R.B.; Music, S.I.; Wenzel, R.; Cash, R.; Libonati, J.P.; Woodward, T.E. 1971. The Broad Street pump revisited: response of volunteers to ingested cholera vibrios. *Bulletin of the New York Academy of Medicine 47(10)*: 1181–1191.

Hoxie, N.J.; Davis, J.P.; Vergeront, J.M.; Nashold, R.D.; Blair, K.A. 1997. Cryptosporidiosis-associated mortality following a massive waterborne outbreak oin Milwaukee, Wisconsin. *American Journal of Public Health87(12)*: 2032–2035.

Ionas, G.; Learmonth, J.J.; Herd, S.D.; Keys, E.A.; Brown, T.J. 1997. What is *Cryptosporidium*? Is it found in New Zealand waters? Can we improve current methods of detection? Proceedings of the Annual Conference of the New Zealand Water and Wastes Association, Rotorua: 115–120.

Johl, M.; Kausch, S.; Eichhorn, D.; Burger, G. 1995. Zur Einschätzung des gesundheitlichen Risikos der Neuen Donau bei deren Nutzung als Badegewässer aus virologischer Sicht. *Forum-Städte-Hygiene 46*: 284–289 (cited by Schernewski & Jülich 2001).

Knight, V. 1964. The use of volunteers in medical virology *in* "Progress in Medical Virology, E. Barger and J.L. Melnic (eds), Based, S. Karger, pp. 2–26.

Koenraad, P.M.F.; Rombouts, F.M.; Notermans, S.H.W. 1997: Epidemiological aspects of thermophilic *Campylobacter* in water-related environments: a review. *Water Environment Research 69(1)*: 52–63.

Landis, J. R.; Koch, G. G. 1977: The measurement of observer agreement for categorical data. *Biometrics 33*: 159–174.

Lee, P.M. 1989. *Bayesian Statistics: An Introduction.* Edward Arnold, London.

LeJeune, J.T.; Besser, T.E.; Merrill, N.L.; Rice, D. H.; Hancock, D. D. 2001: Livestock drinking water microbiology and the factors influencing the quality of drinking water offered to cattle. *Journal of Dairy Science. 84(8)*: 1856–1862.

Lewis, G. 1995. F-specific bacteriophage as an indicator of human wastes in natural waters and sewage effluents in northern New Zealand. *Water Science & Technology 31(5–6)*: 231–234.

Lewis, G.L.; Greening, G. 2000. Analysis for human viruses in water – the when, why and how. *Waters & Wastes in New Zealand 114*: 32–33.

Lightfoot, N.E. 1989. A prospective study of swimming related illness at six freshwater beaches in Southern Ontario. Unpublished Ph.D. thesis, University of Toronto.

Lopez, C.E.; Dykes, A.C.; Juranek, D.D.; Sinclair, S.P.; Conn, J.M.; Christie, R.W. Lippy, E.C.; Schultz, M.G.; Mires, M.H. 1980. Waterborne giardiasis: a community wide outbreak of disease and a high rate of asymptomatic infection. *American Journal of Epidemiology. 112(4)*: 495–507.

López, J.M.; Szewyzk, R. 2000. Estimating the infection risk in recreational waters from the faecal indicator concentration and from the ratio between pathogens and indicators. *Water Research 34(17)*: 4195–4200.

MacKenzie, W.R.; Hoxie, N.J.; Proctor, M.E.; Gradus, M.S.; Blair, K.A.; Peterson, D.E.; Kazmierczak, J.J.; Fox, K.R.; Addias, D.G.; Rose, J.B.; Davis, J.P. 1994. A massive outbreak in Milwaukee of Cryptosporidium infection transmitted through the public water supply. *New England Journal of Medicine 331(3)*: 161–167.

Manly, B.F.J. 1997. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 2nd ed. Chapman & Hall, London. 399 p.

Mara, D. 2000. Cryptosporidium – a reasonable risk? *Water 21* (International Water Association): 34.

McBride, G.B. 1993. Discussion of "Health effects of swimmers and nonpoint sources of contaminated water", by Calderon *et al*. *International Journal of Environmental Health Research 3*: 115–116.

McBride, G.B. 1997. Evaluation of Colilert® procedures using Cohen's "kappa" statistic as the measure of agreement with a standard. Consultancy Report No. SCJ70201. Report to Environmental Diagnostics, Auckland. 19 p. (with permission)

McBride, G.B. 1998. Results of the Trial Sampling Programme. Technical Paper No. 17. Freshwater Microbiological Research Programme. Ministry for the Environment, Wellington, September.

McBride, G.B. in prep. Development of exact MPN tables using occupancy theory. NIWA Technical Report.

McBride, G.B.; Alexander, R.B.; Elliott, S.; Ude Shankar. 2000a. Regional–scale modelling of water quality. *Water & Atmosphere 8(2)*: 29–31.

*McBride, G.B.; Henderson, H.H.; Willms, W. 1998. Analysis of Canadian data on stock response to different drinking-water sources. NIWA (National Institute of Water and Atmospheric Research), Hamilton; Agresearch, Ruakura Research Centre, Hamilton; Agriculture and Agri-Food, Lethbridge, Canada. 28 p.

McBride, G. B.; Loftis, J. C.; Adkins, N. C. 1993. What do significance tests really tell us about the environment? *Environmental Management 17(4)*: 423–432. Errata in *18*: 317.

McBride, G.B.; Salmond, C.E.; Bandaranayake, D.R.; Turner, S.J.; Lewis, G.D.; Till, D.G. 1998. Health effects of marine bathing in New Zealand. *International Journal of Environmental Health Research 8*: 173–189.

*McBride, G.B.; Thorn, C.; Salmond, C. 1996. Feasibility of bathing-health effects study for New Zealand freshwaters. Ministry for the Environment Report No. 202.

McBride, G.B.; Till, D.G.; Ball, A.; Belton, D.; Taylor, K.; Pyle, E.; Steel, K.; Rhodes, D.; Taylor, M.E.U. 2000b. Freshwater Microbiological Research Programme Manual as at 14 February 2000. Report for Ministry for the Environment. 126 p.

McCullagh, P.; Nelder, J.A. 1995. *Generalized Linear Models*.2nd ed. Chapman & Hall, London.

McCullough, N.B.; Eisele, C.W. 1951a. Experimental human salmonellosis. I. Pathogenicity of strains of *Salmonella meleagridis* and *Salmonella anatum* obtained from spray dried whole egg. *Journal of Infectious Diseases 88*: 278–289.

McCullough, N.B.; Eisele, C.W. 1951b. Experimental human salmonellosis. III. Pathogenicity of strains of *Salmonella newport*, *Salmonella derby* and *Salmonella bareilly* obtained from spray dried whole egg. *Journal of Infectious Diseases 89*: 209–213.

Medema, G.J.; Teunis, P.F.M.; Havelaar, A.H.; Haas, C.N. 1996. Assessment of dose-response relationship of *Campylobacter jejuni. International Journal of Food Microbiology 30*: 101–111.

Messner, M.J.; Chappell, C.L.; Okhuysen, P.O. 2001. Risk assessment for *Crypotosporidium*: a hierarchical Bayesian analysis of human dose response data. *Water Research 35(16)*: 3934–3940.

MfE 1998a. Fact Sheet. Freshwater Microbiological Research Programme. Results of the Trial Sampling Programme. Ministry for the Environment, Wellington.

MfE 1998b. Fact Sheet. Freshwater Microbiological Research Programme. Overview of the Programme. Ministry for the Environment, Wellington.

Miles, M.A. 1973. An investigation into the possible relationship between bathing and ear complaints in Taupo. *New Zealand Sanitarian 28(2)*: 75.

Morris, R.D.; Naumova, E.N.; Levin, R.; Munasinghe, R. 1996. Temporal variation in turbidity and diagnosed gastroenteritis in Milwaukee. *American Journal of Public Health 86(2)*: 237–239.

Moss, D.M.; Chappell, C.L.; Okhuysen, P.C.; DuPont, H.L.; Arrowood, M.J.; Hightower, A.W.; Lammie, P.J. 1998. The antibody response to 27-, 17-, and 15-kDa *Cryptosporidium* antigens following experimental infection in humans. *The Journal of Infectious Diseases 178*: 827–833.

Murray, S.; Sinton, L.; Savill, M. 2000. *Enteroviruses* – Indicators of viral pollution of water. *Waters & Wastes in New Zealand 114*: 26–31.

Nagels, J.W.; Davies-Colley, R.J.; Donnison, A.M.; Muirhead, R.W. in press. Faecal contamination over flood events in a pastoral agricultural stream in New Zealand. *Water Science & Technology*.

Okhuysen, P.C.; Chappell, C.L.; Sterling, C.R.; Jakubowski, W.; DuPont, H.L. 1998. Susceptibility and serologic response of healthy adults to reinfection with *Cryptosporidium parvum*. *Infection and Immunity 66(2)*: 441–443.

Okhuysen, P.C.; Chappell, C.L.; Crabb, J.H.; Sterling, C.R.; DuPont, H.L. 1999. Virulence of three distinct *Cryptosporidium parvum* isolates for healthy adults. *The Journal of Infectious Diseases 180*: 1275–1281.

Palisade Corporation. 2000. @RISK. Advanced Risk Analysis for Spreadsheets. Version 4.0.5. Newfield, NY.

Parkhurst, D.F.; Stern, D.A. 1998. Determining average concentrations of *Cryptosporidium* and other pathogens in water. *Environmental Science and Technology 32*: 3424–3429.

Prüss, A. 1998. Review of epidemiological studies on health effects from exposure to recreational water. *International Journal of Epidemiology 27*: 1–9.

Public Health Laboratory Service. 1959. Sewage contamination of coastal bathing waters in England and Wales: a bacteriological and epidemiological study. *Journal of Hygiene (Cambridge) 57*: 435–472. (See also "Sewage contamination of bathing beaches in England and Wales". Medical Research Council Memorandum (UK) No. 37, 32 p.)

Pyle, E. 1995. Principles and processes for the development of guidelines and standards. May.

Regli, S.; Rose, J.B.; Haas, C.N.; Gerba, C.P. 1991. Modeling the risk from Giardia and viruses from drinking water. *Journal of AWWA 83(11)*: 76–84.

Rendtorff, R.C. 1954a. The experimental transmission of human intestinal protozoan parasites. I. *Endamoeba coli* cysts given in capsules. *American Journal of Hygiene 59*: 196–208.

Rendtorff, R.C. 1954b. The experimental transmission of human intestinal protozoan parasites. II. *Giardia lamblia* cysts given in capsules. *American Journal of Hygiene 59*: 209–220.

Rendtorff, R.C.; Holt, C.J. 1954a. The experimental transmission of human intestinal protozoan parasites. III. Attempts to transmit *Endamoeba coli* and *Giardia lamblia* by flies. *American Journal of Hygiene 60*: 320–326.

Rendtorff, R.C.; Holt, C.J. 1954b. The experimental transmission of human intestinal protozoan parasites. IV. Attempts to transmit *Endamoeba coli* and *Giardia lamblia* by water. *American Journal of Hygiene 60*: 327–338.

Rollins, D.M.; Tribble, D.R.; Oplinger, M.L.; Scott, D.A.; Trespalacios, F.; Bourgeois, A.L.; Walz, S.W. 1999. Microbiological outcomes in volunteers receiving experimental oral challenge with live *Campylobacter jejuni*. 10th International Workshop on CHRO (*Campylobacter*, *Helicobacter*, and Related Organisms), 12–16 September 1999, Baltimore MD, USA. Notes p. 107.

Rose, J.B. 1997. Environmental ecology of *Cryptosporidium* and public health implications. *Annual Review of Public Health 18*: 135–161.

Rose, J.B.; Haas, C.N.; Regli, S. 1991. Risk assessment and control of waterborne Giardiasis. *American Journal of Public Health 81(6)*: 709–713.

Rose, J.B.; Gerba, C.P. 1991. Use of risk assessment for development of microbial standards. *Water Science & Technology 24(2)*: 29–34.

Rose, J.B.; Sobsey, M.D. 1993. Quantitative risk assessment for viral contamination of shellfish and coastal waters. *Journal of Food Protection 56(12)*: 1043–1050.

Savill, M.; Hudson, J.A.; Ball, A.; Sholes, P.; Whyte, R.J.; Janovic, D. 2000. Campylobacter in water: enough to make you sick? Proceedings of Water 2000, New Zealand Water & Wastes Assoc., Institution of Professional Engineers of New Zealand, Royal Society of New Zealand, Water Environment Federation (USA), Auckland, March (proceedings available on CD-Rom from NZWWA).

Savill, M.G.; Hudson, J.A.; Ball, A.; Klena, J.D.; Scholes, R.J.; Whyte, R.J.; McCormick, R.E.; Jankovic, D. 2001. Enumeration of *Campylobacter* in New Zealand recreational and drinking waters. *Journal of Applied Microbiology 91*: 38–46.

Scarsbrook, M.R.; Quinn, J.M.; Boothroyd, I.K.G. 2000. Reply: A balanced approach to interpretation of data. *New Zealand Journal of Marine and Freshwater Research 34(4)*: 746–748.

Schernewski, G.; Jülich, W.-D. 2001. Risk assessment of virus infections in the Oder estuary (southern Baltic) on the basis of spatial transport and virus decay simulations. *International Journal of Hygiene and Environmental Health 203*: 317–325.

Schiff, G.M.; Stefanovic, G.M.; Young, E.C.; Sander, D.S.; Pennekamp, J.K.; Ward, R.L. 1984a. Studies of Echovirus-12 in volunteers: determination of minimal infectious dose and the effect of previous infection on infectious dose. *Journal of Infectious Diseases 150(6)*: 858–866.

Schiff, G.M.; Stefanovic, G.; Young, B.; Pennekamp, J.K. 1984b. Minimum human infective dose of enteric virus (Echovirus 12) in drinking water. *Monographs on Virology 15*: 222–228.

Seyfried, P.L.; Tobin, R.S.; Brown, N.E.; Ness, P.F. 1985a. A prospective study of swimming-related illness. I. Swimming-associated health risk. *American Journal of Public Health* **75**, 1068–1070.

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

Simmons, G.; Hope, V.; Lewis, G.; Whitmore, J.; Gao, W. 2001. Contamination of potable roof-collected rainwater in Auckland, New Zealand. *Water Research 35(6)*: 1518–1524.

Sinton, L.W.; Donnison, A.M.; Hastie, C.M. 1993a. Faecal streptococci as faecal pollution indicators: a review. Part I: Taxonomy and enumeration. *New Zealand Journal of Marine and Freshwater Research 27*: 101–115.

Sinton, L.W.; Donnison, A.M.; Hastie, C.M. 1993b. Faecal streptococci as faecal pollution indicators: a review. Part II: Sanitary significance, survival, and use. *New Zealand Journal of Marine and Freshwater Research 27*: 117–137.

Smith, C.M.; Wilcock, R.J.; Vant, W.N.; Smith, D.G.; Cooper, A.B. 1993. Towards sustainable agriculture: freshwater quality in New Zealand and the influence of agriculture. *MAF Policy technical paper 93/10*. Wellington, Ministry of Agriculture and Fisheries. 208 p.

Smith, R.A.; Schwarz, G.E.; Alexander, R.B. 1997. Regional interpretation of water-quality monitoring data. *Water Resources Research 33(12*): 2781–2798.

Sorvillo, F.J.; Fukioka, F.; Nahlen, B.; Tormey, P.; Kebebjian, R.S.; Mascola, L. 1992. Swimming-associated Cryptosporidios. *American Journal of Public Health 82(5)*: 742–744.

SPSS 1998. Systat version 9.01. SPSS Inc., Chicago IL.

Stehr-Green, J.K.; Nicholls, C. McEwan, S.; Payne, A.; Mitchell, P. 1991: Waterborne outbreak of Campylobacter jejuni in Christchurch: the importance of a combined epidemiologic and microbiologic investigation. *New Zealand Medical Journal 104*, 356–358.

Stewart-Tull, D.E.; Coote, J.G.; Thompson, H.; Wardlaw, A.C.; Owen, R.; Bolton, F.; Candlish, A.; Candlish, D. 1999. In vitro and in vivo assays to measure the virulence of *Campylobacter jejuni*. 10th International Workshop on CHRO (*Campylobacter*, *Helicobacter*, and Related Organisms), 12–16 September 1999, Baltimore MD, USA. Notes p. 98.

Suptel, E.A. 1963. Pathenogenesis of experimental Coxsackie virus infection. *Archives of Virology 7*: 61–66.

Teunis, P.F.M.; Havelaar, A.H.; Medema, G.H. 1995. A literature survey on the assessment of microbiological risk for drinking water. Report no. 734301006, RIVM, Bilthoven, The Netherlands

Teunis, P.F.M.; Havelaar, A.H. 1999. *Cryptosporidium* in drinking water: Evaluation of the ILSI/RSI quantitative risk assessment framework. RIVM Report No. 284 550 006, Antonie van Leeuwenhoeklaan 9, PO Box 1, NL-3720 BA Bilthoven, The Netherlands.

Teunis P.F.M.; Havelaar A.H. 2000. The Beta Poisson Dose-Response Model Is Not a Single-Hit Model. *Risk Analysis 20(4)*: 513–520.

Teunis, P.F.M.; van der Heijden, O.G.; van der Giessen, J.W.B.; Havelaar, A.H. 1996. The dose-response relation in human volunteers for gastro-intestinal pathogens. RIVM Report No. 284 550 002, Antonie van Leeuwenhoeklaan 9, PO Box 1, NL-3720 BA Bilthoven, The Netherlands.

Teunis, P.F.M.; Medema, G.J.; Kruidenier, L.; Havelaar, A.H. 1997. Assessment of the risk of infection by *Cryptosporidium* or *Giardia* in drinking water from a surface water source. *Water Research 31(6)*: 1333–1346.

Teunis, P.F.M.; Nagelkerke, N.J.D; Haas, C.N. 1999. Dose Response Models For Infectious Gastroenteritis. *Risk Analysis 19(6)*: 1251–1260.

Teunis, P.F.M.; Chappell, C.L.; Ockhuysen, P.C. 2002a. *Cryptosporidium* dose response studies: variation between isolates. *Risk Analysis 22(1)*: 175–183.

Teunis, P.F.M.; Chappell, C.L.; Ockhuysen, P.C. 2002b. *Cryptosporidium* dose response studies: variation between hosts. *Risk Analysis 22(3)*: 475–485.

Thomas, C.; Gibson, H.; Hill, D.J.; Mabey, M. 1999. *Campylobacter* epidemiology: an aquatic perspective. *Journal of Applied Microbiology Sy,posium Supplement 85*: 168S–177S.

*Till, D.G.; McBride, G.B.; Ball, A.; Taylor, K.; Pyle, E. 2000. Pathogens and Indicators in New Zealand Recreational Freshwaters. Presented at the 10th Health-Related Water Microbiology Symposium, 1st World Water Congress, IWA, Paris, 3–6 July.

*Till, D.G.; McBride, G.B.; Ball, A. 2002. Pathogens and Indictors in New Zealand Freshwaters. IWA Congress, Melbourne, April 2002.

Tillett, H.O.; Coleman, R. 1985. Estimated numbers of bacteria from non-homogeneous bodies of water. *Journal of Applied Bacteriology, 59*: 381–388.

Tillett, H.O.; Sellwood, J.; Lightfoot, N.F.; Boyd, P.; Eaton, S. 2001. Correlations between microbial parameters from water quality samples: expectations and reality. *Water Science & Technology 43(12)*: 19–22.

Tribble, D.R.; Baqar, S.; Oplinger, M.L.; Trespalacios, F.; Scott, D.; Rollins, D.; Walz, S.W.; Burg, E.; Bourgeois, A. 1999. Duration of protection in humans following experimental infection with *Campylobacter jejuni*. 10th International Workshop on CHRO (*Campylobacter*, *Helicobacter*, and Related Organisms), 12–16 September 1999, Baltimore MD, USA. Notes p. 103.

Velleman, P.F. 1997. *DataDesk, version 6.0, Handbook and Statistics Guides*. Data Description Inc., Ithaca, New York.

Ward, R.L.; Akin, E.W. 1984. Minimum infective dose of animal viruses. *CRC Critical Reviews of Environmental Control 14*: 297–310.

Ward, R.L.; Bernstein, D.L.; Young, C.E.; Sherwood, J.R.; Knowlton, D.R.; Schiff, G.M. 1986. Human rotavirus studies in volunteers: determination of infectious dose and serological response to infection. *Journal of Infectious Diseases 154(5)*: 871.

Wiedenmann, A.; Krüger, P.; Gommel, S.; Hirlinger, M.; Eissler, M.; Paul, A.; Jüngst, K.; Botzenhart, K. 2002. An epidemiological study on health risks from bathing in German fresh water bathing sites. Presented at the 11th Symposium of the Health-Related Water Microbiology Group of the International Water Association, Melbourne, Australia. April.

Willms, W.D.; Kenzie, O.R.; McAllister, T.A.; Colwell, D.; Veira, D.; Wilmshurst, J.F.; Entz, T.; Olson, M.E. submitted. Effects of water quality on cattle performance. *Journal of Range Management.*

Willms, W.D.; Kenzie, O.R.; McAllister, T.A.; Colwell, D.; Veira, D.; Wilmshurst, J.F.; Beck, R. 2000. Water quality effects on cattle performance. *In Adams, B. and Douwes, H. (eds), The Range: Progress and Potential*, Proceedings of the Western Range Seminar. Lethbridge, Alberta, Canada.

Wheeler, J.G.; Sethi, D.; Cowden J.M.; Wall, P.G.; Rodrigues, L.C.; Tompkins, D.S.; Hudson, M.J.; Roderick, P.J. 1999. Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. *British Medical Journal 318(7190)*: 1046–1050.

APPENDICES

A.1 Programme Management Group

Chair

Eric Pyle, Ministry for the Environment, Wellington.

Members

Desmond Till, Consultant Microbiologist, Wellington.

Ken Taylor, Senior Analyst, Environment Canterbury, Christchurch.

Phil Journeaux/David Rhodes/Kevin Steele, Ministry of Agriculture and Forestry. Wellington

Graham McBride, Principal Scientist, NIWA (National Institute of Water and Atmospheric Research) , Hamilton.

Andrew Ball Microbiologist, ESR (Institute of Environmental Science and Research), Christchurch.

Dr Derek Belton, Food Programme Leader, ESR (Institute of Environmental Science and Research), Mt. Albert, Auckland—now MAF Policy, Wellington

Dr Michael Taylor, Senior Advisor Environment, Ministry of Health, Wellington

Assisting staff: Megan Linwood (Ministry for the Environment, Wellington)
Ruth Berry (Ministry for the Environment, Wellington)

A.2 Risk Analysis Working Group

Chair

Desmond Till, Consultant Microbiologist, Wellington.

Members

Dr Gillian Lewis, Senior Lecturer Microbiology and Virology Research Group, School of Biological Sciences, University of Auckland.

Associate Professor Philip Weinstein, Senior Lecturer, Department of Public Health, Wellington School of Medicine, Otago University.

Andrew Ball Microbiologist, ESR (Institute of Environmental Science and Research), Christchurch.

Dr Stephen Palmer, Medical Officer of Health, Hutt Valley Health,

Dr Terry Ryan, Manager (Epidemiology Services), Animal Health Board, Hamilton.

Graham McBride, Principal Scientist, NIWA (National Institute of Water and Atmospheric Research), Hamilton.

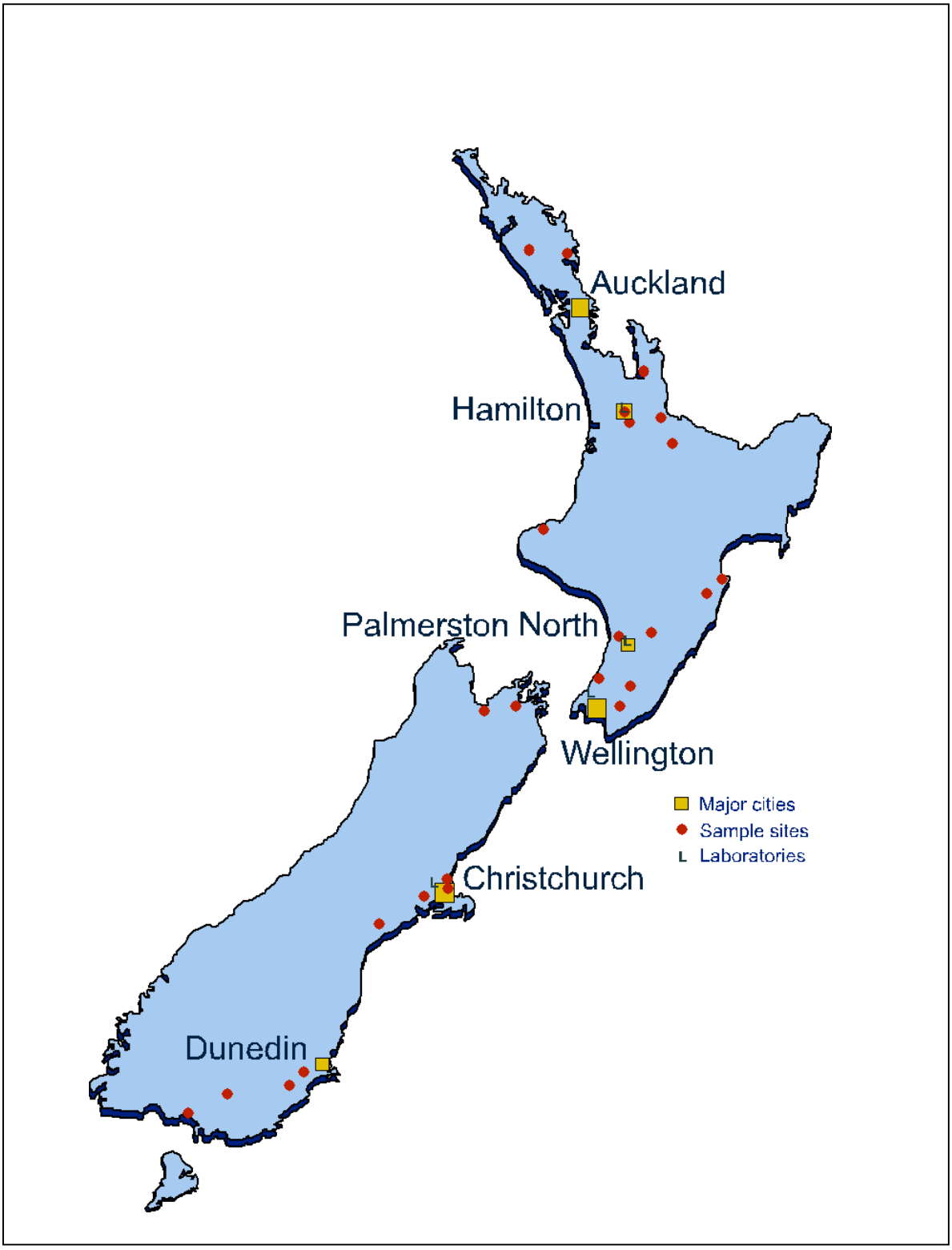
A.3 Main survey analysis

A.3.1 Site locations and categories

Zone	Regional Councils	Sites[†]	Map ref. (NZMS260)	Map ref. (full co-ordinates)	Water quality*	Predominant source(s)
Northern North Island	Northland	Hatea River (Whangarei Falls)	Q06:316-121	2631600 6612100	Marginal	Urban, forest
		Mangakahia River (Twin Bridges)	PO6:874-194	2587400 6619400	Dirty	Forest
	Waikato	Lake Rotoroa (Hamilton)	S14:109-756	2710900 6375600	Dirty	SW/birds
		Kauaeranga River (Smiths)	T12:396-462	2739600 6446200	Marginal	Forest
		Waikato River[#&] (Hamilton water intake)	S14:130-746	2713000 6374600	Marginal	Agricultural, SW
Central North Island	Bay of Plenty	Wairoa River (McLaren Falls)	U14:783-729	2778300 6372900	Dirty	Forestry, dairy, sheep
		Lake Okareka	U16:038-324	2803800 6332400	Clean	Sheep, settlement
	Taranaki	Waiwhakaiho River (S.E.M. site)[#]	P19:084-287	2608400 6228700	Moderate	Dairying, water supply
	Hawke's Bay	Tukituki River (Black Bridge)[#]	V21:477-702	2847700 6170200	Moderate	Sheep, urban, sewage
		Tukituki River (Waipukurau)[#]	V23:136-291	2813600 6129050	Moderate	Sheep
Lower North Island	Wellington	Otaki River (quarry, upstream of cableway)	S25:955-405	2695500 6040500	Clean	Forest
	Wairarapa	Ruamahanga River (Double Bridges)[&]	T26:344-335	2734400 6033500	Moderate	Forest, beef, sheep
		Ruamahanga River (Morrisons Bush)	S27:188-017	2718800 6001700	Clean	Dairy, sheep
	Manawatu	Oroua River (Feilding)	S23:297-047	2729700 6104700	Dirty	Sheep, beef, dairy
		Pohangina River (Raumai)	T23: 474-072	2747400 6107200	Moderate	Forest, pastoral
Upper South Island	Tasman	Lee River (Picnic Area, Valley Road)	N28:222-775	2522200 5977500	Moderate	Forestry – exotic & native
	Marlborough	Onamalutu River (at Scenic Reserve—Ford)	O28:690-721	2569009 5972118	Moderate	Forest
	Canterbury	Waimakariri River (upstream of SH1 Bridge)[&]	M35:813-546	2481280 5754630	Marginal	Extensive pastoral, birds
		Selwyn River (Coes Ford)	M36:626-234	2462600 5723400	Dirty	Dairying, sheep
		Avon River (Antigua Boatsheds)	M35:799-414	2479900 5741400	Dirty	Birds/SW
Ashburton		Ashburton River (Ollivers Rd.)[#]	K37:019-088	2401914 5708784	Dirty	Extensive pastoral, birds
Lower South Island	Southland	Oreti River (Iron Bridge)[#]	E46:454-208	2145400 5420800	Dirty	Sheep, beef, dairy, water supply
		Mataura River (Gore)	F45:967-487	2196700 5448700	Dirty	Sheep, beef, dairy
	Otago	Taieri River (Outram)	I44:955-803	2295500 5480300	Moderate	Extensive pastoral
		Lake Waihola (Waihola)	H45:849-613	2284900 5461300	Moderate	Dairy, sewage, sheep

[†] All but two are recreational sites: the Waiwhakaiho River site (a Taranaki Regional Council State-of-the-Environment Monitoring "S.E.M" site) is just above a recreational site; the Ashburton River site is strictly for water supply only. [*] Based on water quality records supplied by Regional Councils. [#] Also source waters for community supply. "SW" = stormwater. [&]Ruamahunga site is the same as that used in the preliminary study while the Waikato and Waimakariri sites are upstream of the sites of the same name used in the used in the preliminary study.

Sample sites and laboratory locations



A.3.2 Site groups

After consultations with Regional Council staff, the following site groups have been agreed. The site category (e.g., "Birds") reflects the expected dominant influencing factor on the microbiological quality of their waters. The detail of this classification was not revealed to those performing statistical analyses until late in that analysis.

B: **Birds**
- MA Lake Rotoroa (Hamilton)
- EC Avon River, Antigua Boatsheds
- WA Ashburton River, Ollivers Road
- CA Waimakariri River, just upstream of SH bridge

D: **Dairy**
- OC Waikato River, Hamilton water intake
- HA Waiwhakaiho River, Taranaki Regional Council S.E.M. site
- RB Ruamahanga River, Morrison's Bush
- SA Oroua River, Feilding
- DB Selwyn River, Coes Ford

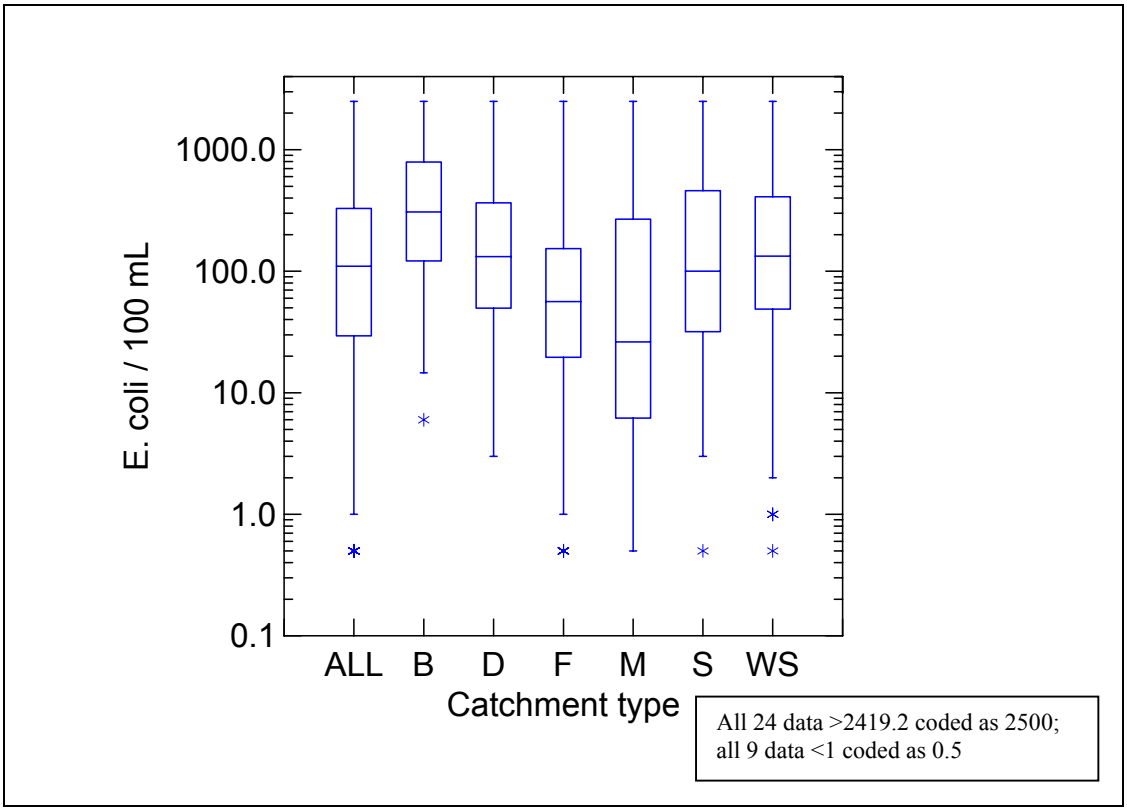
F: **Forestry/undeveloped**
- LB Mangakahia River, Twin Bridges
- FA Wairoa River, McLaren Falls
- PA Otaki River, by quarry, upstream of cableway
- TB Pohangina River, Raumai
- AA Lee River, Picnic area, Valley Road
- NB Kauaeranga River, Smiths
- BA Onamalutu River, Scenic reserve (Ford)

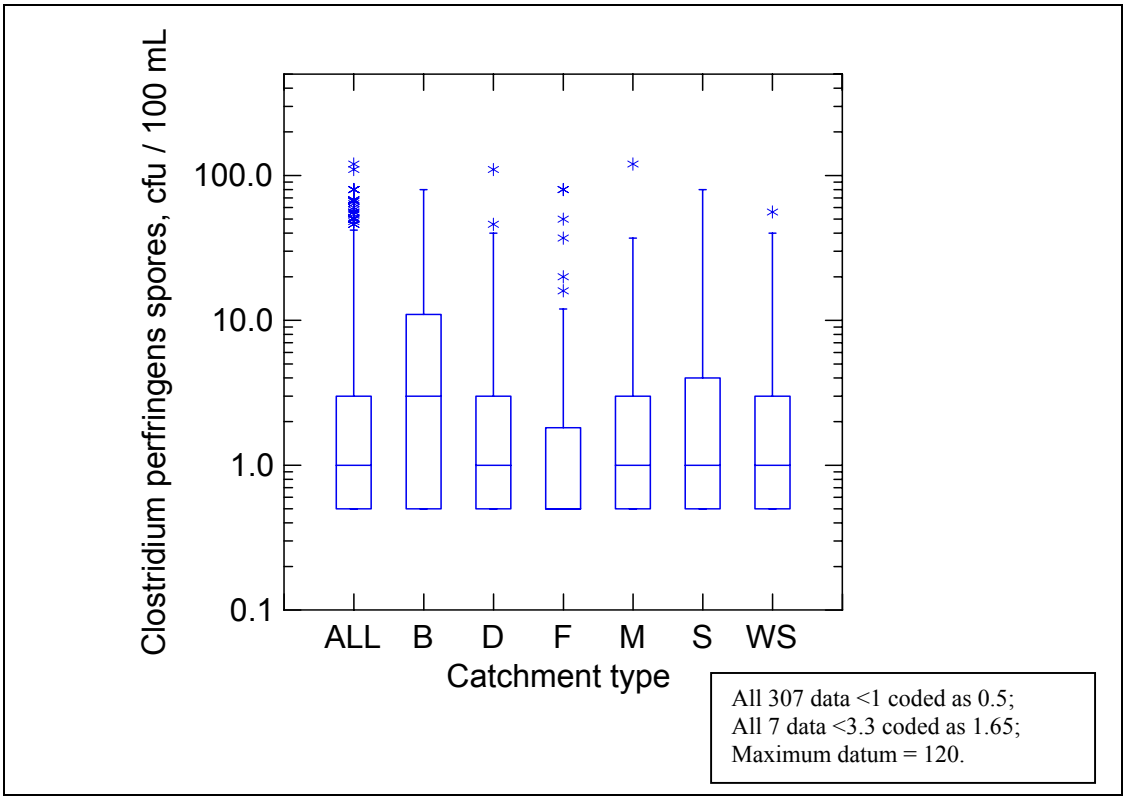
M: **Municipal**
- KA Hatea River, Whangarei Falls
- GB Lake Okareka
- IA Tukituki River, Black Bridge

S: **Sheep/pastoral**
- JB Tukituki River, Waipukurau
- QA Ruamahanga River, Double Bridges
- XA Taieri River, Outram
- UA Oreti River, Iron Bridge
- VB Mataura River, Gore
- YB Lake Waihola, at Waihola

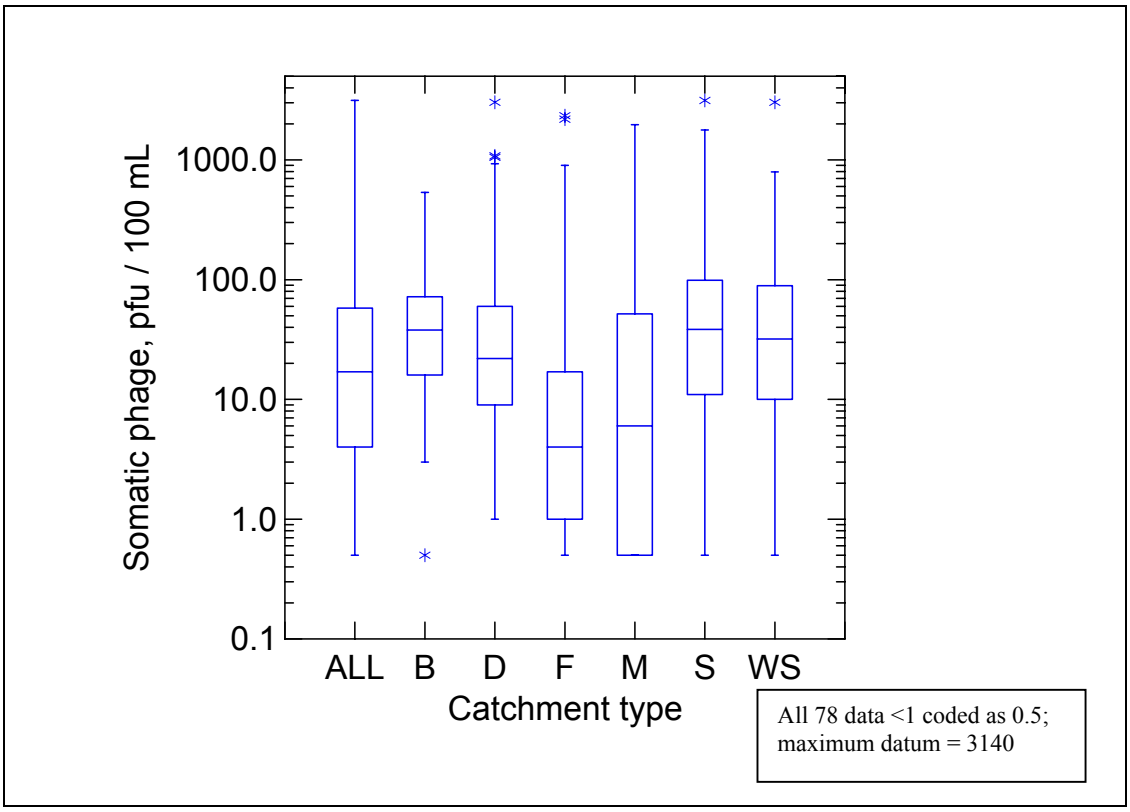
WS: **Water Supply**
- HA Waiwhakaiho River, Taranaki Regional Council S.E.M. site
- IA Tukituki River, Black Bridge
- JB Tukituki River, Waipukurau
- WA Ashburton River, Ollivers Road
- UA Oreti River, Iron Bridge

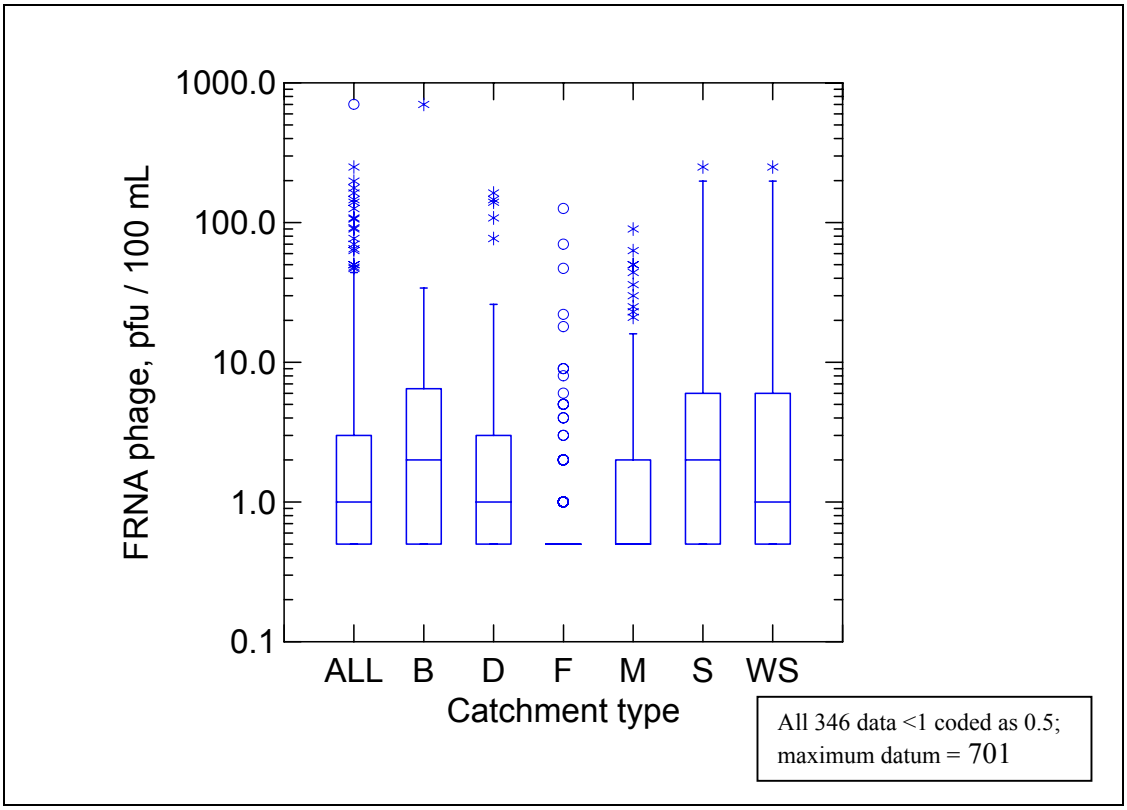
A.3.3 Spatial variability of analytes over all site groups



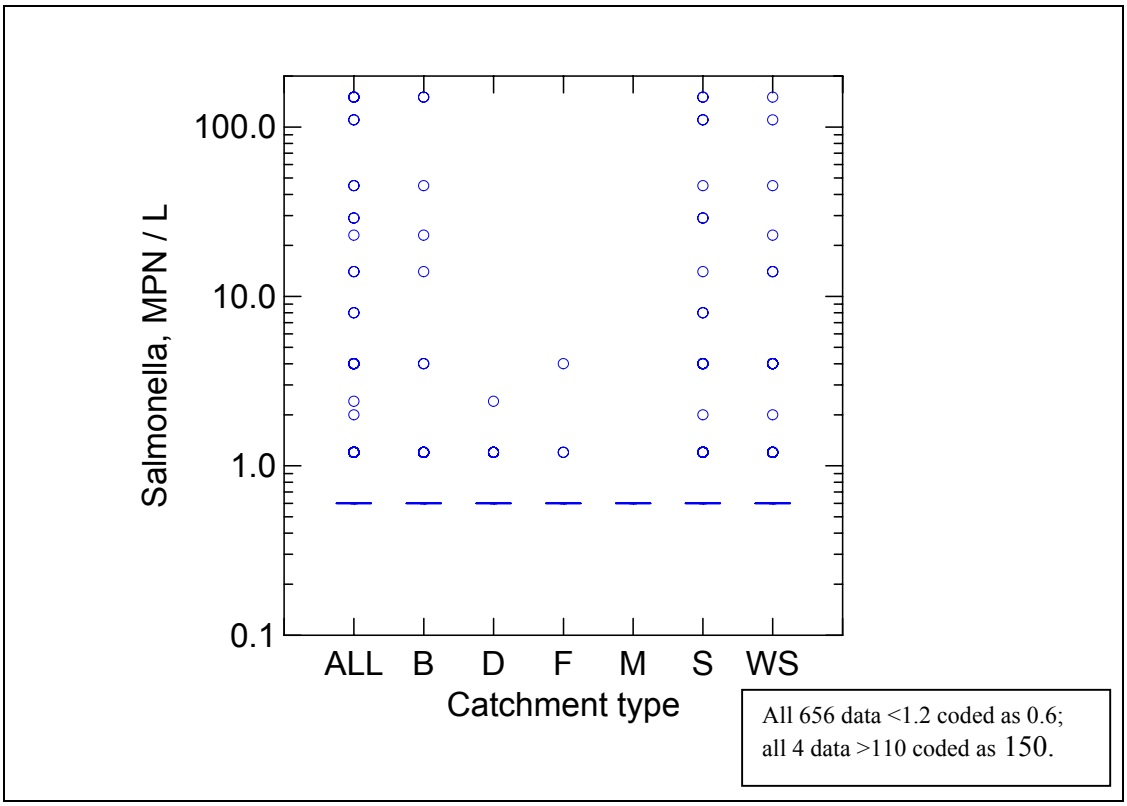


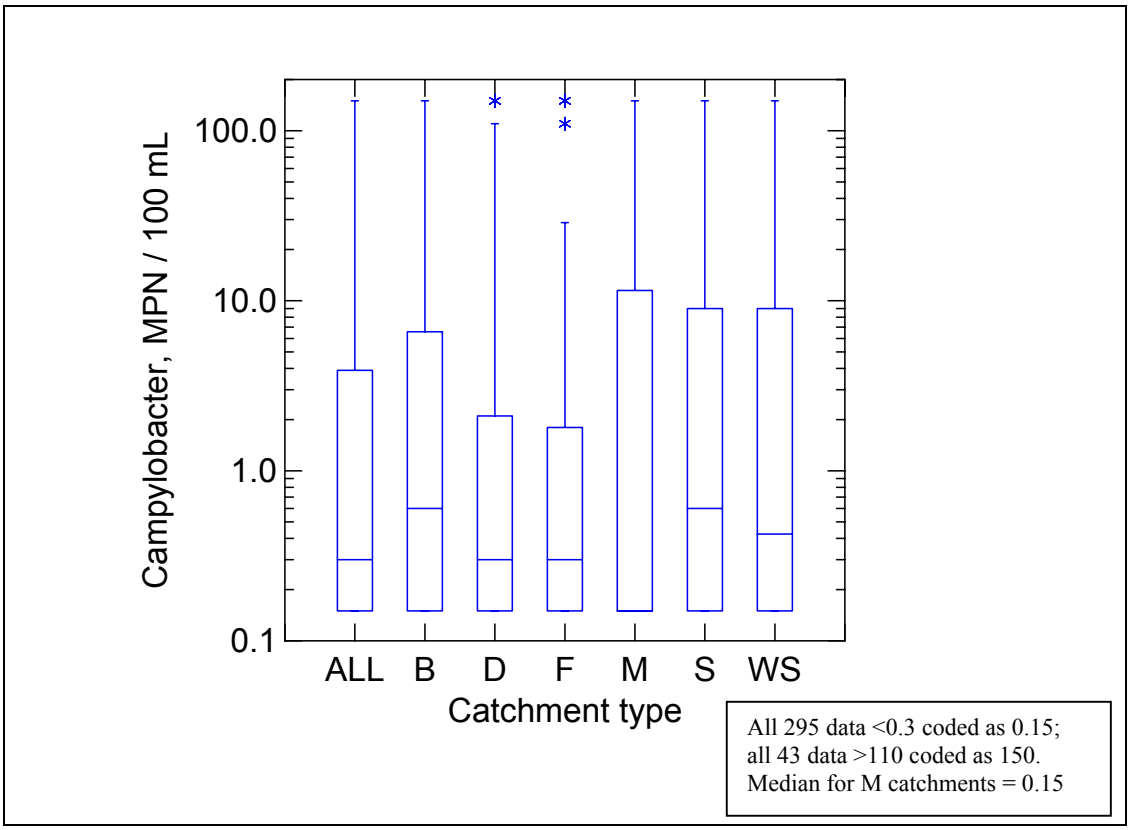
Boxplots for *E. coli* and *C. perfringens* spores. These were prepared using Systat (SPSS 1998). The box covers the interquartile range (IQR); the bottom of the box is the 25%ile and the top is the 75%ile. The line within the box is the median (50%ile). Whiskers extend from the box to the furthest datum inside the inner fence (defined as 1.5xIQR beyond the box). Data between the inner and outer fence (defined as 3xIQR beyond the box) are outliers and are plotted as a starburst. Data beyond the outer fence (e.g., in the Salmonella data) are extreme outliers and are plotted as circles.



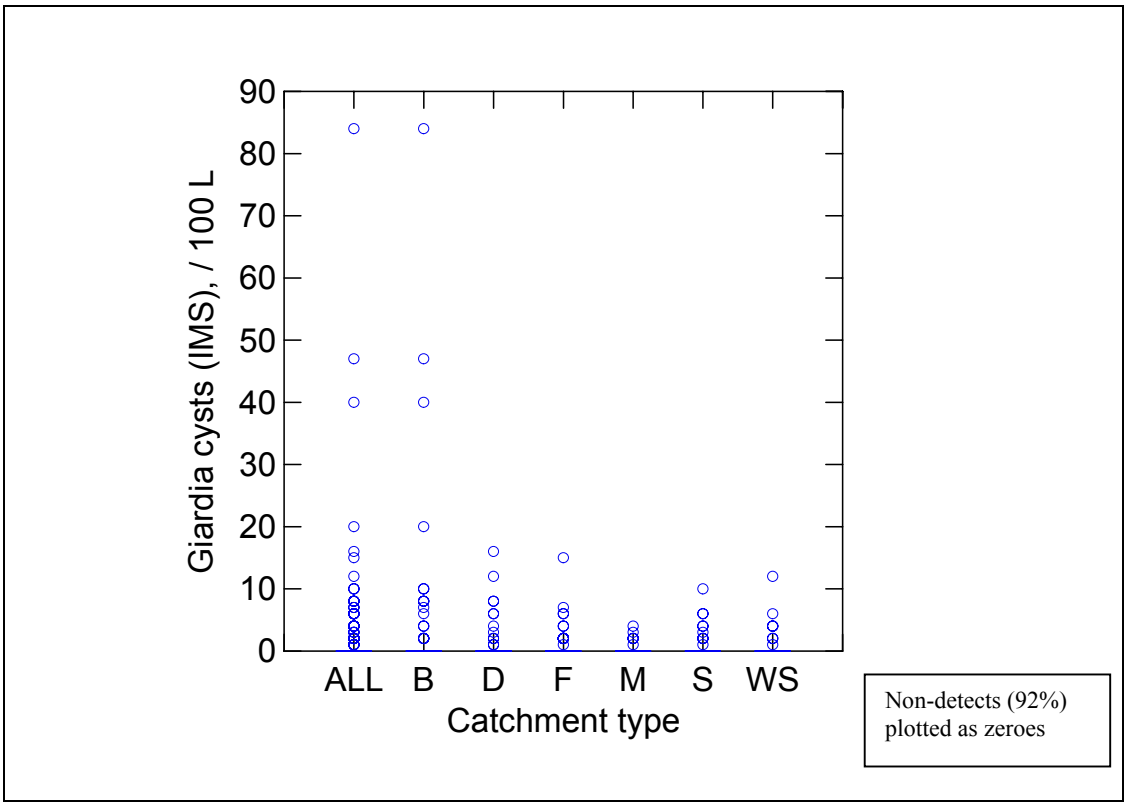


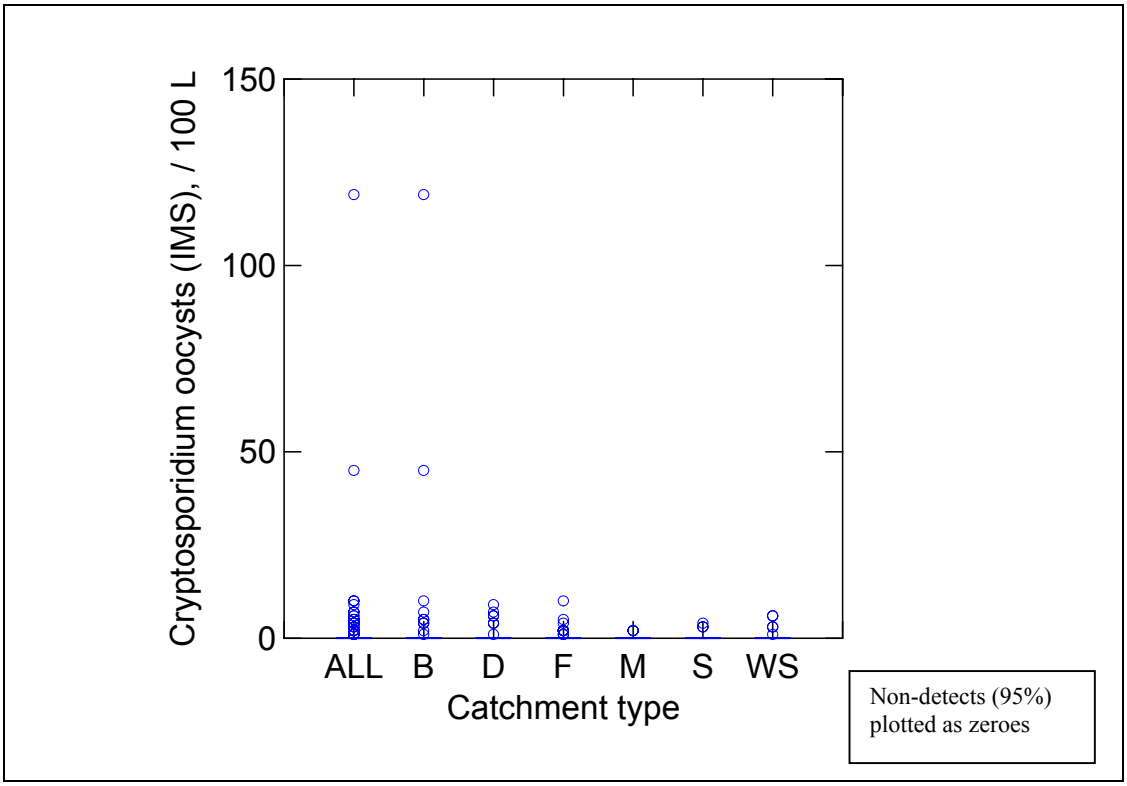
Boxplots for somatic coliphage and FRNA phage. These were prepared using Systat (SPSS 1998). The box covers the interquartile range (IQR); the bottom of the box is the 25%ile and the top is the 75%ile. The line within the box is the median (50%ile). Whiskers extend from the box to the furthest datum inside the inner fence (defined as 1.5xIQR beyond the box). Data between the inner and outer fence (defined as 3xIQR beyond the box) are outliers and are plotted as a starburst. Data beyond the outer fence (e.g., in the Salmonella data) are extreme outliers and are plotted as circles.





Boxplots for *Salmonella* and *Campylobacter*. These were prepared using Systat (SPSS 1998). The box covers the interquartile range (IQR); the bottom of the box is the 25%ile and the top is the 75%ile. The line within the box is the median (50%ile). Whiskers extend from the box to the furthest datum inside the inner fence (defined as 1.5xIQR). Data between the inner and outer fence (defined as 3xIQR) are outliers and are plotted as a starburst. Data beyond the outer fence (e.g., in the Salmonella data) are extreme outliers and are plotted as circles.





Boxplots for *Giardia* cysts and *Cryptosporidium* oocysts. These were prepared using Systat (SPSS 1998). The box covers the interquartile range (IQR); the bottom of the box is the 25%ile and the top is the 75%ile. The line within the box is the median (50%ile). Whiskers extend from the box to the furthest datum inside the inner fence (defined as 1.5xIQR beyond the box). Data between the inner and outer fence (defined as 3xIQR beyond the box) are outliers and are plotted as a starburst. Data beyond the outer fence (e.g., in the Salmonella data) are extreme outliers and are plotted as circles.

Enterovirus

Group	No. positive	Total number	percent positive
All	**241**	**725**	**33**
B	42	116	36
D	55	145	38
F	59	203	29
M	30	87	34
S	55	174	32
WS	48	145	33

NB. There are two missing values, on 27/9/99 at sites UA and VB (the bacteria tests for the sampling on 6/9/99 were omitted and so a repeat bacteria-only run was performed on 27/9/99).

Adenovirus

Group	No. positive	Total number	percent positive
All	**233**	**725**	**32**
B	51	116	44
D	31	145	21
F	63	203	31
M	24	87	28
S	64	174	37
WS	61	145	42

NB. There are two missing values, on 27/9/99 at sites UA and VB (the bacteria tests for the sampling on 6/9/99 were omitted and so a repeat bacteria-only run was performed on 27/9/99).

***Giardia* cysts (APHA/IMS)**

Group	No. positive	Total number	percent positive
All	**59**	**725**	**8**
B	17	116	15
D	12	145	8
F	14	203	7
M	6	87	7
S	10	174	6
WS	8	145	6

***Cryptosporidium* oocysts (APHA/IMS)**

Group	No. positive	Total number	percent positive
All	**33**	**725**	**5**
B	10	116	9
D	7	145	5
F	10	203	5
M	3	87	3
S	3	174	2
WS	5	145	3

Means and standard deviations of logarithms (base 10) of FMRP data.

		E. coli		*C. perfringens* spores		Somatic coliphage		FRNA phage		*Campylobacter*	
Group	*n*	$\overline{X}$	*S*	$\overline{X}$	*S*	$\overline{X}$	*S*	$\overline{X}$	*S*	$\overline{X}$	*S*
B	116	2.44	0.56	0.52	0.72	1.54	0.52	0.34	0.60	0.14	0.99
D	145	2.11	0.64	0.14	0.52	1.38	0.65	0.21	0.63	-0.14	0.88
F	203	1.71	0.79	0.005	0.45	0.68	0.85	-0.14	0.39	-0.22	0.85
M	87	1.62	0.98	0.11	0.55	0.78	1.02	0.15	0.63	0.028	1.09
S	174	2.04	0.75	0.23	0.58	1.51	0.72	0.36	0.65	0.15	1.05
All	725	1.97	0.79	0.18	0.58	1.17	0.85	0.16	0.61	-0.03	0.97
WS	145	2.05	0.74	0.11	0.50	1.38	0.77	0.35	0.67	0.10	1.02

López & Szewyk (2000) state that standard deviation of logarithms of *E. coli* (base 10) for rivers in Germany average 0.76. These results are in harmony with that figure. In contrast, note that the lakes results of Dufour (1984) give about 0.4 for this value (as also stated in New Zealand's now-repealed 1992 Provisional Water Quality Guidelines—Department of Health 1992).

Proportion of high *Campylobacter* results among the site types.

Site type	No. of samples	>100 / 100mL
B	116	11 (9.5%)
D	145	9 (6.2%)
F	203	12 (5.9%)
M	87	8 (9.2%)
S	176	24 (13.6%)

Number in parentheses is the proportion of samples in each category.

Table A.3.3.1 *Campylobacter* detections by catchment type and by species

	All	**B**	**D**	**F**	**M**	**S**	**WS**
% detected in all samples	59.5 432/726	71.6 83/116	57.9 84/145	52.5 106/202	49.4 43/87	65.9 116/176	62.8 91/145
% positive samples containing *C. jejuni*[†]	47.9 207/432	51.8 43/83	54.8 46/84	41.5 44/106	58.1 25/43	42.2 49/116	53.9 49/91
% positive samples containing *C. lari*[†]	19.0 82/432	22.9 19/83	11.9 10/84	9.4 10/106	11.6 5/43	32.8[‡] 38/116	24.2 22/91
% positive samples containing *C. coli*[†,#]	1.4 6/432	0 0/83	2.4 2/84	0 0/106	9.3 4/43	0.9 1/116	2.2 2/91
% positive samples containing unidentified thermotolerant species[3]	41.0 177/432	34.9 29/83	34.5 29/84	52.8 56/106	34.9 15/43	41.4 48/116	35.2 32/91

[†] Some samples were detected with more than one species (5 had *C. jejuni* and *C. coli*; 34 had *C. jejuni* and *C. lari*; 1 had *C. jejuni*, *C. coli* and *C. lari*).

[‡] cf. 14% for all other catchment types (i.e., 44/316)

[#] *C. coli* detected alone in only one sample.

[¶] These unidentified species include *C. upsaliensis* and at least some of the other thermotolerant species (*C. lari*, *C. jejuni*, *C. coli*) missed by the testing procedure. To see why, careful note must be taken of the nature of the speciation methodology. The main *Campylobacter* PCR assay detected either *C. lari* (82 detects) or "other thermotolerant species" (350 detects). Formally, these "other" species comprise *C. jejuni*, *C. coli* and *C. upsaleinsis* but the test used cannot identify which of them may have been present (more than one can be present). This "other" category may also include low numbers of *C. lari* that the main assay failed to pick up. Immediately after performing the tests broths from all positive tubes in all positive tests were stored and frozen for possible follow-up speciation analyses. Funding was later obtained for this and during 2001 one broth from the largest positive tube for each of these tests was selected randomly and reanalysed with another PCR test that does identify *C. jejuni* and *C. coli* (but not *C. upsaliensis* or *C. lari*). Accordingly, 173 samples that were identified as "other thermotolerant species" in the main assay were found on the retest to contain either *C. jejuni* and/or *C. coli*. Of the remaining 177 samples in that category some 100 were found to contain unidentified thermotolerant species with the balance (77) indicating "not detected"—presumably reflecting sample deterioration. Accordingly this number (177) contains unknown proportions of all four thermotolerant *Campylobacter* species.

Table A.3.3.2 Percentiles of the *E. coli* distributions

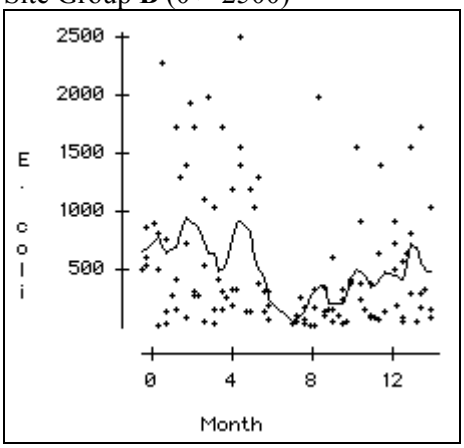
%ile	A	A(bs)	B	D	F	M	S
5	4	5	32	11	2	1	7
10	9	10	48	18	6	2	12
15	14	17	72	27	8	4	20
20	32	25	95	35	14	5	27
25	29	32	122	49	20	6	32
30	40	49	139	57	23	8	41
35	51	62	154	75	31	11	52
40	66	86	179	99	41	12	66
45	91	107	269	117	47	16	83
50	110	133	308	132	56	26	100
55	131	156	326	142	71	80	116
60	154	199	379	160	94	167	140
65	191	258	488	212	110	223	175
70	261	308	613	291	127	245	270
75	332	435	793	371	154	272	461
80	461	548	958	461	179	364	603
85	613	770	1213	548	261	424	816
90	980	1203	1539	1046	416	745	1024
95	1986	1986	1733	1993	1200	1435	2500

Percentiles calculated using DataDesk®. "**A(bs)**" denotes all data in the bathing season (1 November to 31 March). Water supply site category not included. Shaded percentiles correspond to the percentiles of campylobacteriosis infection rates on Table A3.7.3, where those rates first rise above 10/1,000 persons.

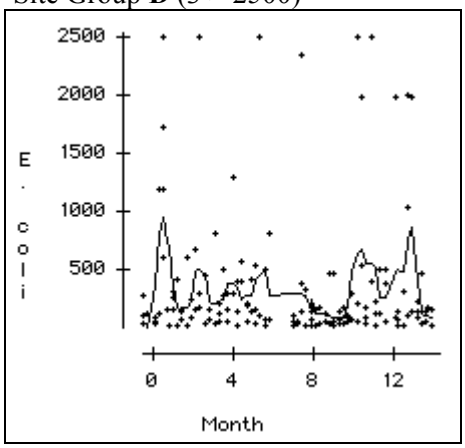
A.3.4 Temporal variability for each analytes over all site groups

E. Coli results (/100 mL) for site groups: **B** (birds), **D** (dairy), **F** (forestry/undeveloped), **M** (municipal), **S** (sheep/pastoral)[41]

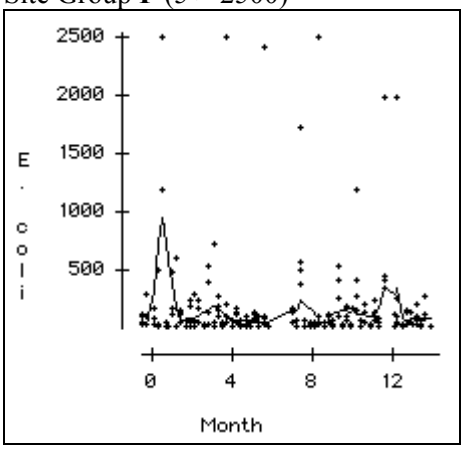
Site Group **B** (0 > 2500)



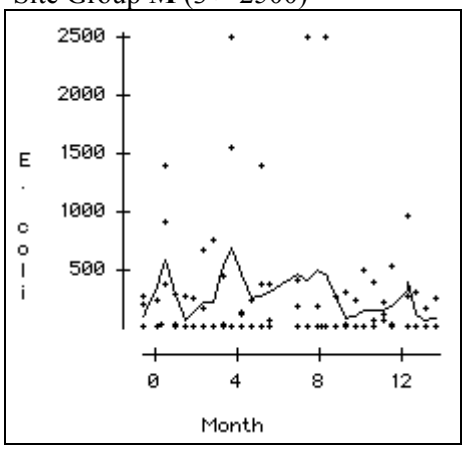
Site Group **D** (3 > 2500)



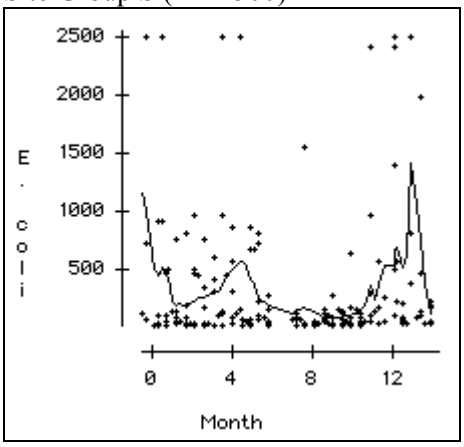
Site Group **F** (5 > 2500)



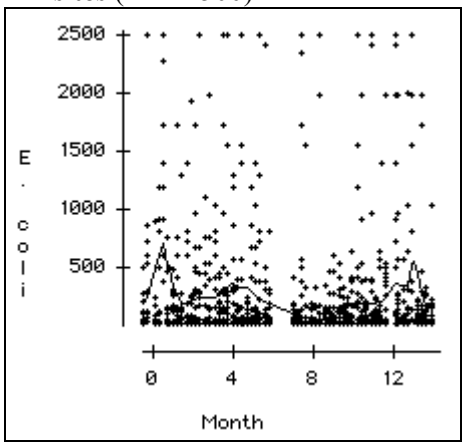
Site Group **M** (3 > 2500)



Site Group **S** (1 > 2500)



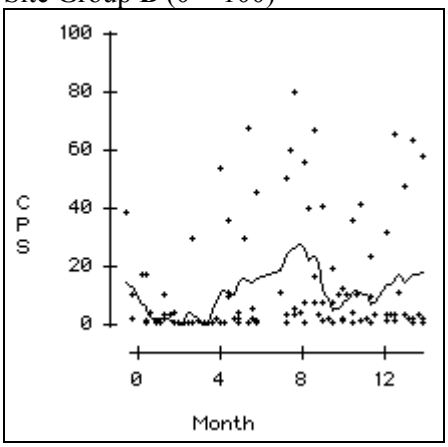
All sites (12 > 2500)



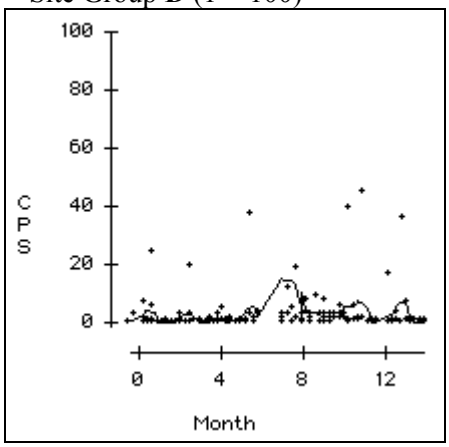
[41] All 9 data "<1" have been set to 0.5; all 12 data ">2419.2" have been reset to 2500, and all 12 data recorded as a value above 2500 have been reset to 2500 also (some early samples were diluted 10x before analysis, so their upper detection limit was >24192"). While it would be useful to portray confidence or credibility intervals for the TREWESS curves, we lack the appropriate software.

Clostridium perfringens spore results (cfu/100 mL) for site groups: **B** (birds), **D** (dairy), **F** (forestry/undeveloped), **M** (municipal), **S** (sheep/pastoral)[42]

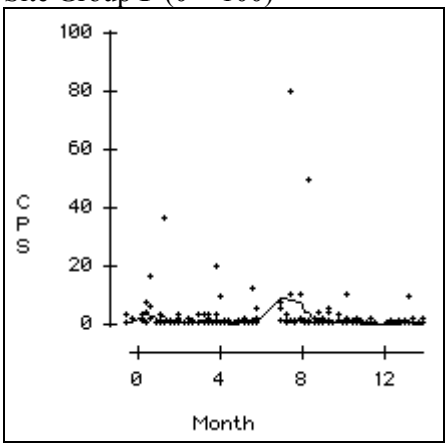
Site Group **B** (0 > 100)



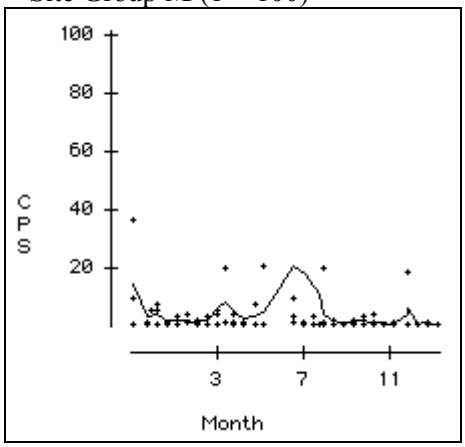
Site Group **D** (1 > 100)



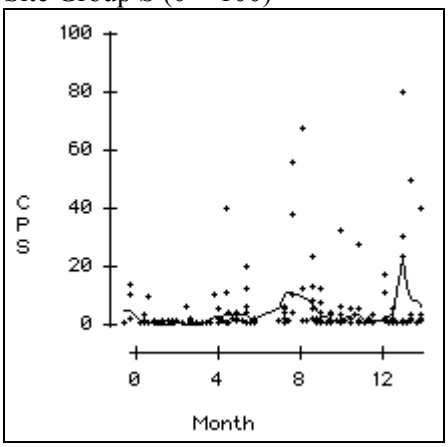
Site Group **F** (0 > 100)



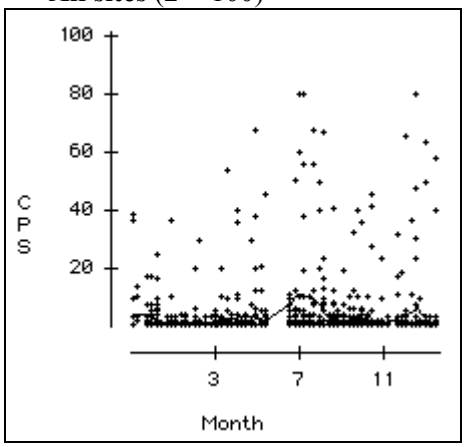
Site Group **M** (1 > 100)



Site Group **S** (0 > 100)



All sites (2 > 100)

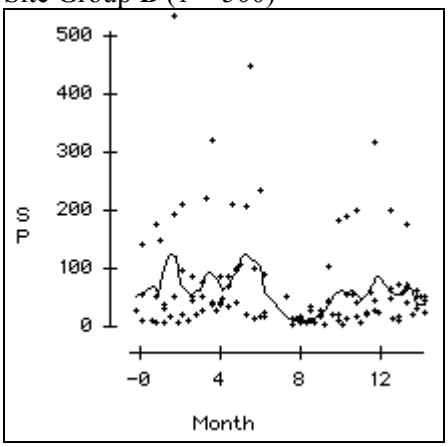


[42] All 307 data "<1" have been replaced by 0.5 and all 7 data "<3.3" have been replaced by 1.65. The following 2 data exceed the graphs' upper limit (/100 mL):

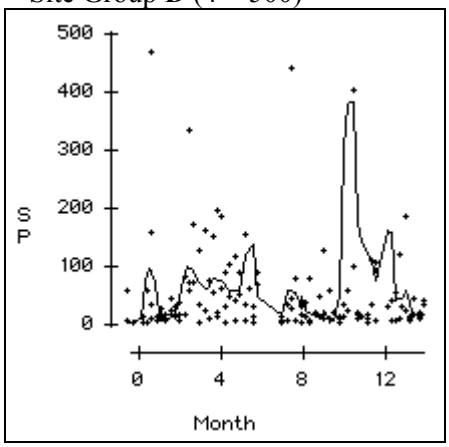
Value	Site	Date	Value	Site	Date
110	SA	16/8/99	120	KA	16/8/99

Somatic coliphage results (pfu/100 mL) for site groups: **B** (birds), **D** (dairy), **F** (forestry/undeveloped), **M** (municipal), **S** (sheep/pastoral)[43]

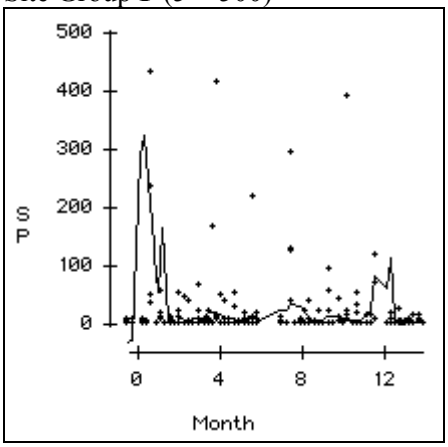
Site Group **B** (1 > 500)



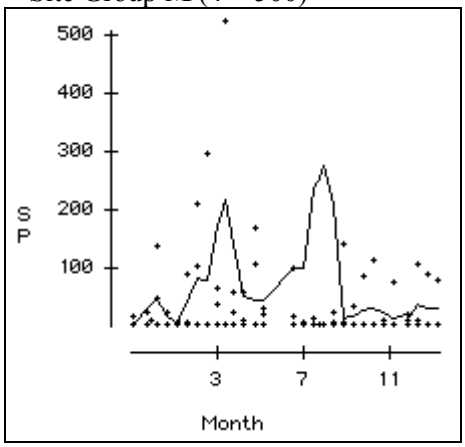
Site Group **D** (4 > 500)



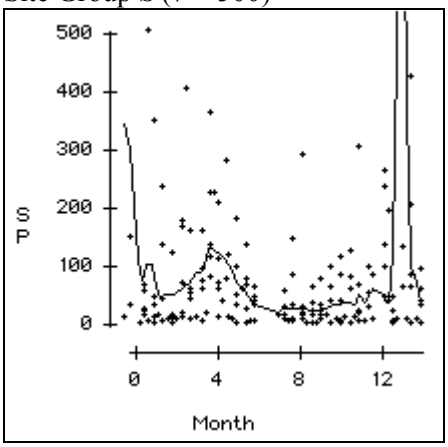
Site Group **F** (5 > 500)



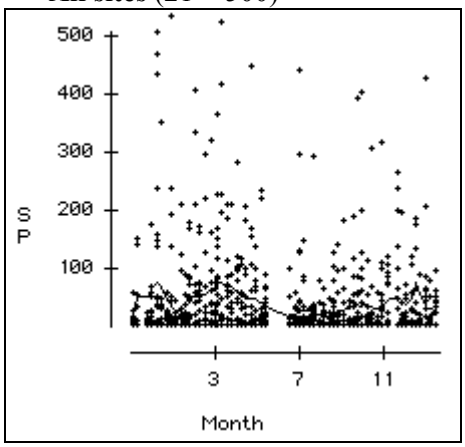
Site Group **M** (4 > 500)



Site Group **S** (7 > 500)



All sites (21 > 500)



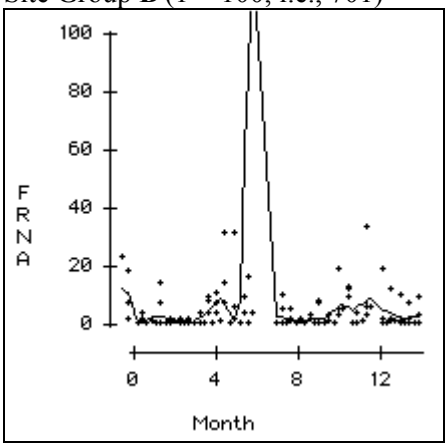
All sites (bottom portion of previous graph)

[43]All 78 data "<1" have been replaced by 0.5. These 21 data exceed the graphs' upper limit (/500 mL):

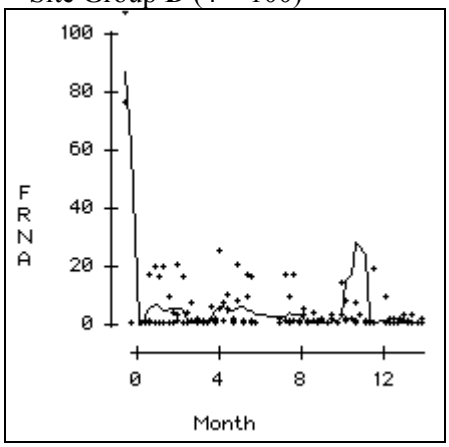
Value	Site	Date	Value	Site	Date	Value	Site	Date	Value	Site	Date	Value	Site	Date
508	JB	18/1/99	659	TB	20/12/99	885	LB	10/1/00	1780	XA	31/1/00	3140	YB	31/1/00
526	IA	26/4/99	722	UA	17/5/99	900	LB	13/9/99	1970	KA	13/9/99			
536	EC	8/2/99	768	KA	26/4/99	930	DB	14/6/99	2190	BA	11/1/99			
568	VB	31/1/00	774	VB	21/12/98	1050	RB	29/11/99	2340	BA	8/2/99			
655	KA	16/8/99	795	UA	21/12/98	1083	DB	5/1/00	3030	HA	8/11/99			

FRNA phage results (pfu/100 mL) for site groups: **B** (birds), **D** (dairy), **F** (forestry/undeveloped), **M** (municipal), **S** (sheep/pastoral)[44]

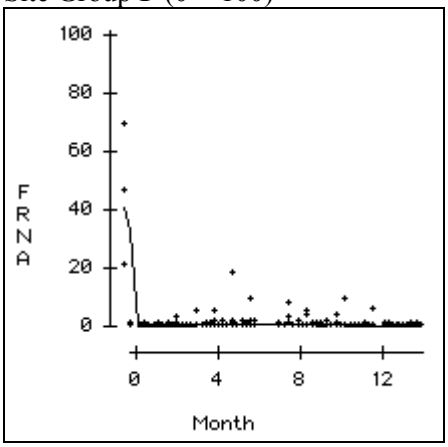
Site Group **B** (1 > 100, i.e., 701)



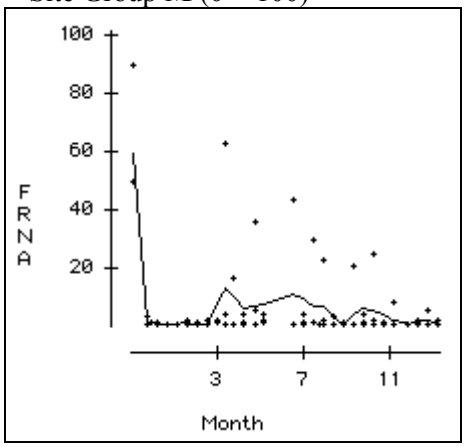
Site Group **D** (4 > 100)



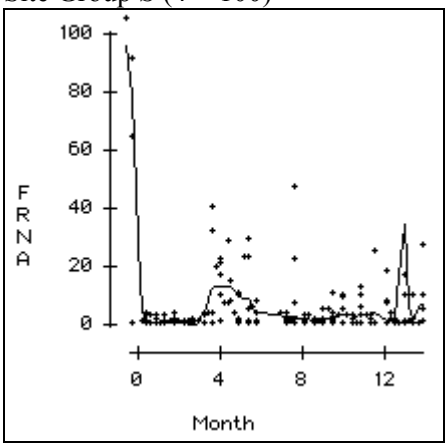
Site Group **F** (0 > 100)



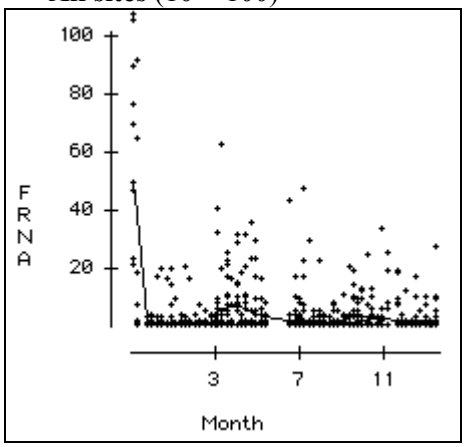
Site Group **M** (0 > 100)



Site Group **S** (4 > 100)



All sites (10 > 100)

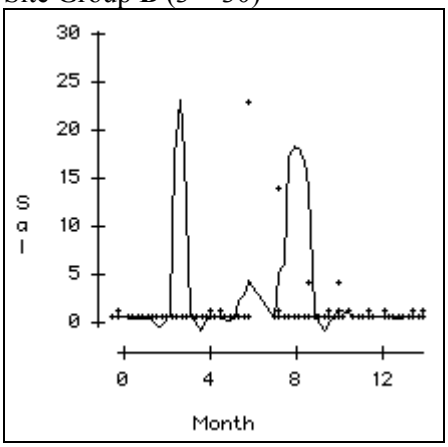


[44] All 346 data "<1" have been replaced by 0.5. The following 10 data exceed the graphs' upper limits:

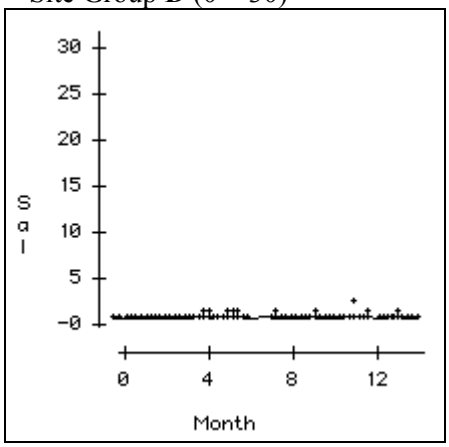
Value	Site	Date	Value	Site	Date	Value	Site	Date	Value	Site	Date
106	JB	14/12/98	140	RB	15/11/99	177	XA	31/1/00	701	EC	28/6/99
108	SA	14/12/98	146	HA	14/12/99	198	UA	21/12/98			
126	TB	14/12/98	163	RB	29/11/99	250	UA	17/5/99			

Salmonella results (MPN/L) for site groups: **B** (birds), **D** (dairy), **F** (forestry/undeveloped), **M** (municipal), **S** (sheep/pastoral)[45]

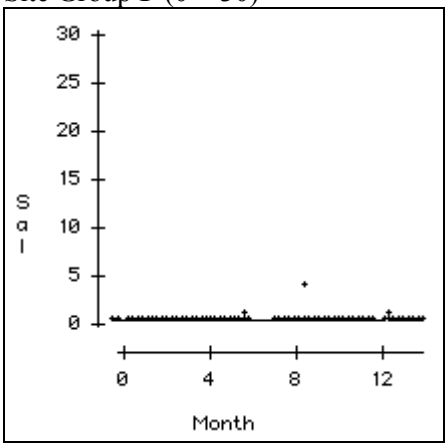
Site Group **B** (3 > 30)



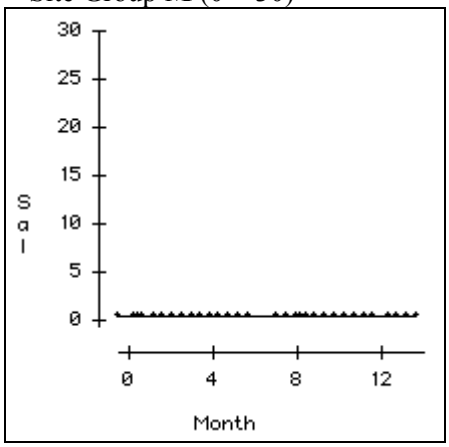
Site Group **D** (0 > 30)



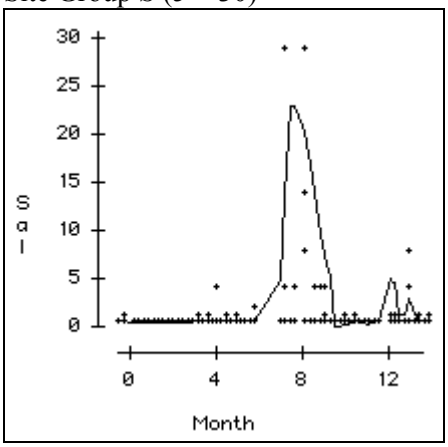
Site Group **F** (0 > 30)



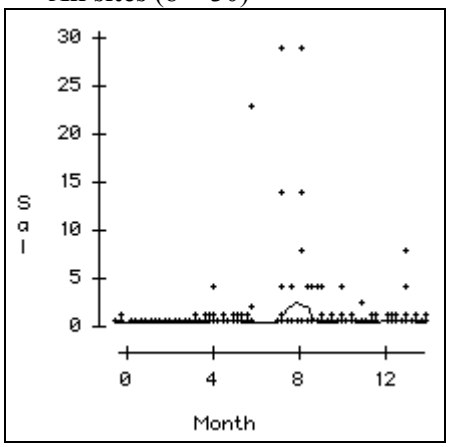
Site Group **M** (0 > 30)



Site Group **S** (5 > 30)



All sites (8 > 30)

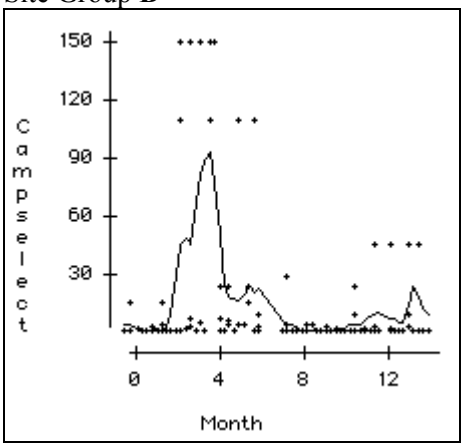


[45] All 656 data "<1.2" have been replaced by 0.6. The following 8 data exceed the graphs' upper limit (30 MPN/L):

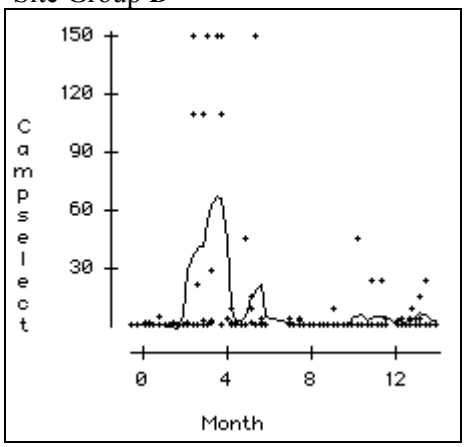
Value	Site	Date	Value	Site	Date	Value	Site	Date	Value	Site	Date
45	WA	23/8/99	110	UA	23/8/99	>110	EC	22/3/99	>110	VB	20/9/99
45	XA	5/1/00	110	VB	27/9/99	>110	VB	23/8/99	>110	WA	6/9/99

Campylobacter results (MPN/100 mL) for site groups: **B** (birds), **D** (dairy), **F** (forestry/undeveloped), **M** (municipal), **S** (sheep/pastoral)[46]

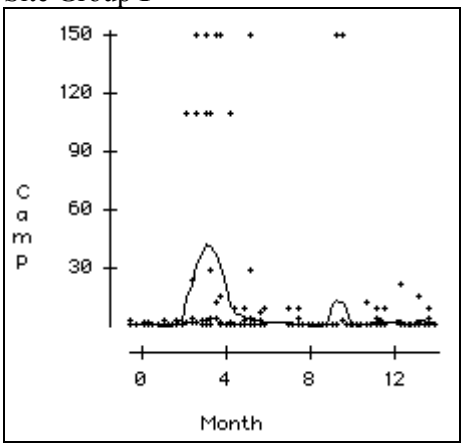
Site Group **B**



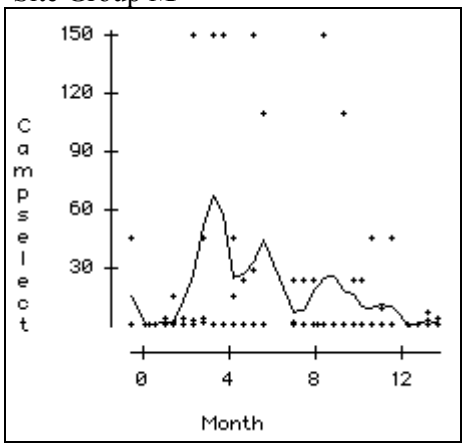
Site Group **D**



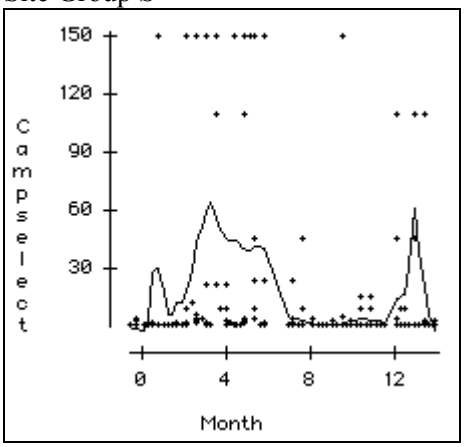
Site Group **F**



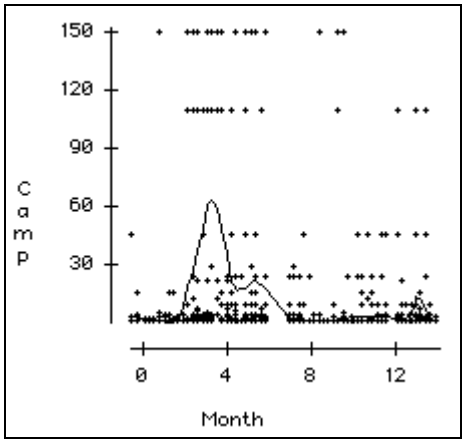
Site Group **M**



Site Group **S**



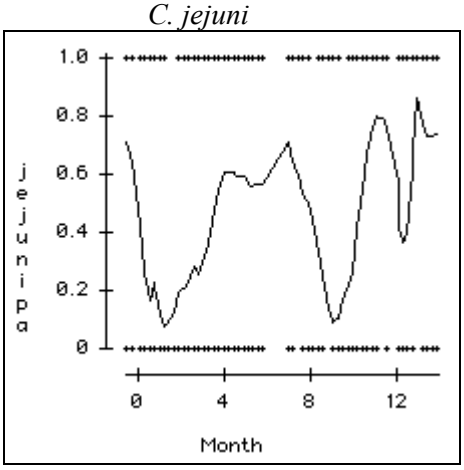
All sites



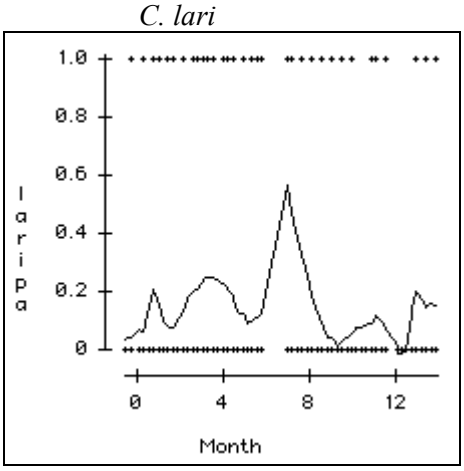
[46] All 295 data "<0.3" have been replaced by 0.15; all 43 data ">110" have been replaced by 150 and so are all shown on these graphs.

Presence/absence trends for *C. jejuni*, *C. lari*, Human Adenovirus, Human Enterovirus, *Giardia* cysts and *Cryptosporidium* oocysts

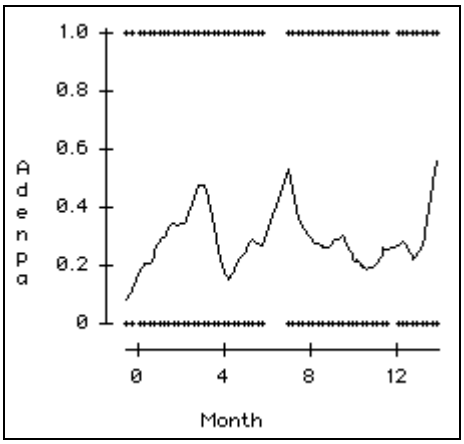
C. jejuni



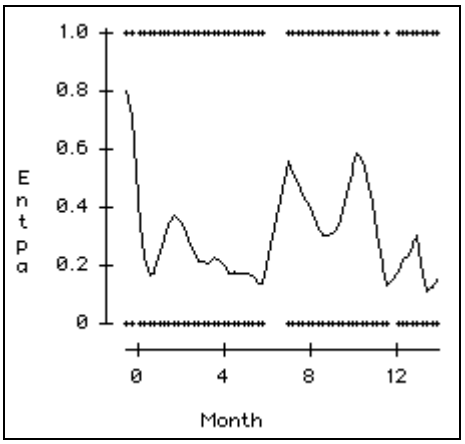
C. lari



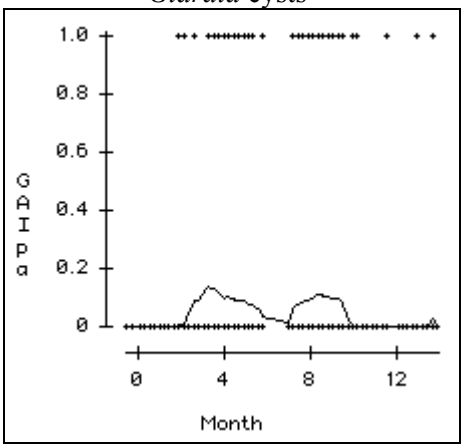
Adenovirus



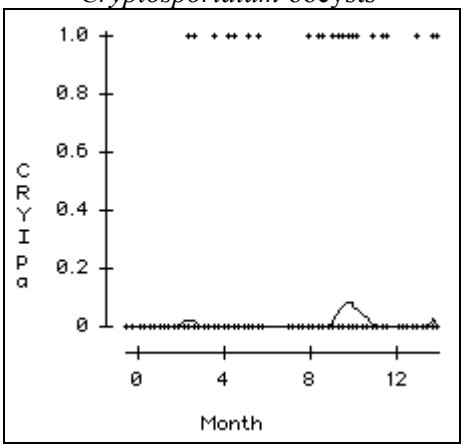
Enterovirus



Giardia cysts



Cryptosporidium oocysts



A.3.5 Correlations between analytes over all site groups

A.3.5.1 Overall correlation coefficients, excluding virus data[47,48]

Pearson Product-Moment Correlation ("*r*")

	EC	CPS	SP	FRNA	Sal	Camp	GAI	CRYI
EC	1.000							
CPS	0.457	1.000						
SP	0.551	0.335	1.000					
FRNA	0.189	0.217	0.171	1.000				
Sal	0.084	0.129	0.013	0.020	1.000			
Camp	0.260	0.042	0.210	0.063	0.004	1.000		
GAI	0.068	0.296	0.057	0.583	0.010	0.006	1.000	
CRYI	0.158	0.131	0.029	0.001	0.151	-0.013	0.004	1.000

Spearman Rank Correlation ("r_s")

	EC	CPS	SP	FRNA	Sal	Camp	GAI	CRYI
EC	1.000							
CPS	0.353	1.000						
SP	0.715	0.335	1.000					
FRNA	0.392	0.276	0.471	1.000				
Sal	0.214	0.262	0.278	0.243	1.000			
Camp	0.416	0.094	0.401	0.246	0.166	1.000		
GAI	0.047	0.169	0.072	0.119	0.092	0.119	1.000	
CRYI	0.060	0.053	0.014	0.016	0.023	0.064	0.084	1.000

A.3.5.2 Spearman Rank Correlations, bathing season only[49]

	EC	CPS	SP	FRNA	Sal	Camp	GAI	CRYI
EC	1.000							
CPS	0.421	1.000						
SP	0.728	0.381	1.000					
FRNA	0.379	0.273	0.466	1.000				
Sal	0.308	0.275	0.332	0.286	1.000			
Camp	0.352	0.128	0.368	0.230	0.205	1.000		
GAI	-0.038	0.031	0.035	0.038	-0.003	0.040	1.000	
CRYI	0.042	0.037	0.030	0.035	0.080	0.108	0.155	1.000

[47] Spearman's rank correlation coefficient (r_s also called Spearman's rho) is computed using the same formula as one does for the more common (Pearson's) *r*, except that it uses the data's ranks rather than their actual values. Pearson's *r* measures the degree of *linear* correlation whereas Spearman's r_s measures the degree of monotonic correlation and so is more general. [If a relationship were exponentially increasing (or decreasing) Spearman's coefficient would be very close to +1 (or –1), whereas *r* may be much closer to zero, giving a false impression about the lack of correlation]. Sometimes *r* can exceed r_s, but not often. One needs to guard against findings of "no correlation" when either coefficient is in some sense "low", especially with the small samples sizes typical of microbiological studies (Tillett *et al.* 2001).

[48] *E. coli* (EC), *C. perfingens spores* (CPS), somatic phage (SP), FRNA phage (FRNA), *Salmonella* (Sal), *Campylobacter* (Camp), Giardia by APHA/IMS (GAI), Cryptosporidium by APHA/IMS (CRYI)

[49] Defined here and elsewhere as 1 November to 31 March.

A.3.5.3 Spearman (rank) correlations, by site group

B (bird) catchments

	EC	CPS	SP	FRNA	Sal	Camp	GAI	CRYI
EC	1.000							
CPS	0.301	1.000						
SP	0.514	0.036	1.000					
FRNA	0.196	0.083	0.204	1.000				
Sal	0.110	0.143	0.094	0.151	1.000			
Camp	0.360	0.011	0.298	0.228	0.182	1.000		
GAI	0.035	0.289	0.064	0.190	-0.019	0.107	1.000	
CRYI	0.112	0.174	-0.004	0.059	0.101	0.089	0.047	1.000

D (dairying) catchments

	EC	CPS	SP	FRNA	Sal	Camp	GAI	CRYI
EC	1.000							
CPS	0.186	1.000						
SP	0.718	0.182	1.000					
FRNA	0.322	0.165	0.343	1.000				
Sal	0.201	0.256	0.181	0.154	1.000			
Camp	0.343	-0.033	0.360	0.114	0.129	1.000		
GAI	0.030	0.202	0.102	0.166	0.115	0.101	1.000	
CRYI	0.052	0.055	0.065	0.062	-0.061	0.195	0.176	1.000

F catchments (Forestry/undeveloped)

	EC	CPS	SP	FRNA	Sal	Camp	GAI	CRYI
EC	1.000							
CPS	0.380	1.000						
SP	0.724	0.334	1.000					
FRNA	0.235	0.148	0.343	1.000				
Sal	0.151	0.061	0.169	0.234	1.000			
Camp	0.255	0.019	0.159	0.102	0.012	1.000		
GAI	0.016	0.136	0.056	0.242	-0.033	0.182	1.000	
CRYI	0.092	-0.010	0.036	-0.002	0.171	0.021	0.029	1.000

M catchments (municipal)[50]

	EC	CPS	SP	FRNA	Sal	Camp	GAI	CRYI
EC	1.000							
CPS	0.446	1.000						
SP	0.768	0.436	1.000					
FRNA	0.488	0.372	0.516	1.000				
Sal	•	•	•	•	•			
Camp	0.541	0.328	0.595	0.352	•	1.000		
GAI	0.080	0.024	-0.014	-0.103	•	0.048	1.000	
CRYI	0.018	-0.056	-0.084	-0.043	•	0.036	0.194	1.000

[50] The dots reflect the fact that no positive results were obtained for Salmonella in Municipal catchments.

S catchments (sheep)

	EC	CPS	SP	FRNA	Sal	Camp	GAI	CRYI
EC	1.000							
CPS	0.155	1.000						
SP	0.594	0.265	1.000					
FRNA	0.469	0.340	0.430	1.000				
Sal	0.223	0.380	0.330	0.212	1.000			
Camp	0.507	0.064	0.541	0.325	0.187	1.000		
GAI	-0.019	0.128	0.045	0.032	0.248	0.070	1.000	
CRYI	-0.080	0.042	-0.052	-0.017	-0.066	-0.024	-0.033	1.000

WS water supply sites

	EC	CPS	SP	FRNA	Sal	Camp	GAI	CRYI
EC	1.000							
CPS	0.419	1.000						
SP	0.876	0.425	1.000					
FRNA	0.504	0.373	0.433	1.000				
Sal	0.256	0.437	0.257	0.209	1.000			
Camp	0.566	0.183	0.551	0.267	0.205	1.000		
GAI	0.085	0.164	0.103	0.034	0.290	0.163	1.000	
CRYI	0.064	0.007	0.090	0	-0.084	0.132	0.131	1.000

A.3.5.3 Associations between presence/absence data

The following table gives values for "Cohen's kappa" for all comparisons of presence/absence data over the site groups and for all sites. This statistic is a general-purpose measure of inter-rater agreement, whereby two rates give dichotomous (presence/absence) assessments of a quantity. It can be generalized to a measure of association between two variables, as shown below. The cyst and oocyst data are included in this table, as present/absent data.

Comparison	**B** (*n*=116)[†]	**D** (*n*=145)	**F** (*n*=203)	**M** (*n*=87)	**S** (*n*=174)	**All** (*n*=725)	**WS** (*n*=145)
Aden/Ent	0.019[‡] 19,32,23,42[§]	0.104 15,16,40,74	–0.007 18,45,41,99	–0.175 5,19,25,38	0.121 25,39,30,80	0.029 82,151,159,333	0.082 23,38,25,59
Aden/CRYI	0.023 5,46,5,60	–0.086 0,31,7,107	–0.003 3,60,7,133	–0.065 0,24,3,60	–0.034 0,64,3,107	–0.021 8,225,25,467	–0.068 0,61,5,79
Aden/GAI	–0.056 6,45,11,54	0.023 3,28,9,105	0.019 5,58,9,131	–0.049 1,23,5,58	0.010 4,60,6,104	0.000 19,214,40,452	0.052 5,56,3,81
Ent/CRYI	0.017 4,38,6,68	–0.058 1,54,6,84	0.098 6,53,4,140	–0.002 1,29,2,55	0.002 1,54,2,117	–0.041 13,228,20,464	–0.067 0,48,5,92
Ent/GAI	0.036 7,35,10,64	–0.054 3,52,9,81	–0.002 4,55,10,134	–0.130 0,30,6,51	–0.005 3,52,7,112	–0.020 17,224,42,442	0.014 3,45,5,92
CRYI/GAI[¶]	0.044 2,8,15,91	0.159 2,5,10,128	0.027 1,9,13,180	0.185 1,2,5,79	–0.027 0,3,10,161	0.077 6,27,53,639	0.116 1,4,7,133

[†]*n* is the number of samples in each comparison. [‡]Cohen's kappa ($\kappa = 1$: perfect agreement; $\kappa = 0$: no agreement beyond that expected by chance alone; $\kappa = -1$: complete discordance). [§]Numbers of present/present, present/absent, absent/present and absent/absent data for each comparison. [¶] Comparing the 237 samples that have cysts measured by two methods (APHA and IMS) we find values of Cohen's kappa of –0.036 for *Cryptosporidium* oocysts and 0.061 for *Giardia* cysts (the present/present, present/absent, absent/present and absent/absent data were 0, 13, 6 and 218 for oocysts, and 1, 5, 14 and 217 for cysts).

Discussion: Cohen's kappa

This draws on material in two standard works (Bishop *et al*. 1975; Fleiss 1981), as summarised in a client report by McBride 1997 (cited with permission). Consider the frequency table below:

Frequency table

	Enterovirus		
Adenovirus	Presence	Absence	Total
Presence	a	b	p_1
Absence	c	d	q_1
Total	p_2	q_2	1

All the cell entries are frequencies (i.e., counts divided by n). The simplest measure of agreement is the "overall proportion of agreement", i.e., $p_o = a + d$. However this ignores the agreement that could have arisen merely by chance. Cohen (1960) introduced a measure (κ, the Greek letter kappa) that does account for this, for which one needs to also consider the chance-expected table.

Chance-expected table

	Enterovirus		
Adenovirus	Presence	Absence	Total
Presence	p_1p_2	p_1q_2	p_1
Absence	q_1p_2	q_1q_2	q_1
Total	p_2	q_2	1

From the table the "overall proportion of chance-expected agreement" is $p_e = p_1p_2 + q_1q_2$. The estimated value of Cohen's kappa is then simply given by

$$\hat{\kappa} = \frac{p_o - p_e}{1 - p_e}$$

where the "hat" ($\hat{}$) denotes an estimate (of the unknown true value, κ).[51] A value $\hat{\kappa} = 0$ indicates no agreement beyond that due to chance alone; $\hat{\kappa} = 1$ indicates perfect agreement. Values less than zero indicate that the observed agreement is less than that attributable to chance alone.

What does the kappa statistic tell us?

As a rough guide to intermediate values Landis and Koch (1977) suggested:

Kappa Statistic	*Strength of Agreement*
< 0.00	Poor
0.00–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost Perfect

Values of $\hat{\kappa}$ in this study are so low that none of the associations measured are anything more than "slight".

[51] One can show algebraically that kappa may also be estimated by $2(ad–bc)/(p_1q_2+ p_2q_1)$, giving the same result.

A.3.6 General linear modelling

Extensive general linear modelling has been performed of each of the indicators/pathogens within each of the five site types, including data on river flows and/or turbidity, to see if these ancillary variables play an important role in explaining determinand variability, using the following scheme:

Stream flow, lake level	**Analysis objectives**	
	Bathing season data only (November – March)	All data
Low flows only (<25%ile) & lakes during dry periods	• Health effects on recreational users	• Baseline
All flows and levels	• Effects of higher flows • Resident microbial populations	• Environmental impacts (catchment type) • Drinking water

The following tables give *p*-values for various models of determinands versus a number of combinations of environmental factors, using:

- analysis of variance (ANOVA) where the dependent variables (*E. coli*, clostridia, somatic phage, FRNA phage and *Campylobacter*) are continuous—"interval scale"[52]—but all the environmental factors are categorical (i.e., not continuous)
- analysis of covariance (ANCOVA) where the dependent variable is continuous and at least one of the environmental factors is continuous also (i.e., not categorical).
- logistic models, where the dependent variables (i.e., viruses and cysts) are dichotomous (present/absent) and environmental factors are either continuous or categorical.

This modelling has been performed using DataDesk (Velleman 1997), using "Type 3" sums of squares (also called "partial sums of squares").[53] Each box on the tables contains *p*-values for the stated model (those in brackets refer to the model with no interaction terms).

In interpreting results one must bear in mind that *p*-values[54] should be used for *comparison* of effects for a given sample size; they are not to be used as an *absolute* measure of the strength of effects. This is because the hypothesis tested, while convenient, is never believable—that there is actually no effect whatsoever, down to the zillionth decimal place (Berger & Sellke 1987, Lee 1989; McBride *et al*. 1993; Goodman 1993, 1999). So we are looking for *patterns* in the *p*-values, without taking too much notice of their individual values. For example, consider the following page—for *E. coli* models. The *p*-values for the last two boxes in the left-hand column show that either flow (on the day of sampling, or in the preceding days) or turbidity are important explicatory variables (because their *p*-values are so low compared to *p*-values for other variables). The southeast box in the tables contains the final model, and it shows that Catchment type is also an important explicatory variable.

[52] Strictly, the reported *Campylobacter* concentration is not a continuous variable, because its laboratory test provides only a limited number of "Most Probable Numbers" that it can attain. Nevertheless it has been treated as though it is continuous.

[53] Partial sums of squares offer tests of effects of each factor *after removing the linear effects of all the other factors*. The alternative is "Type I" sums of squares ("sequential sums of squares"). These remove the effects of factors in sequential order, testing the effects of each after the previous terms have been removed. Quite different results can occur between these two options. We chose to partial sums of squares because there is no reason requiring that factors be removed in a specified order.

[54] *p*-values are defined as the probability of getting data at least as large if in fact the factor has no effect at all. They tend to get bigger as the number of samples increases, so care must be taken in comparing their values in the different boxes on the tables. They are properly used to compare relative strength of effects for a given sample size (*n*), not as an absolute measure of strength.

Models for *E. coli*

	Bathing season only		**All times**	
Low flows	Model: $\log_{10}(EC) = f(Ctc)$		Model: $\log_{10}(EC) = f(Ctc, BSn)$	
	n	176	*n*	241
	Const	≤0.0001	Const	≤0.0001 (≤0.0001)
	Ctc	≤0.0001	*Ctc*	≤0.0001 (≤0.0001)
			BSn	0.0131 (0.0018)
			*Ctc*BSn*	0.1289
	Model: $\log_{10}(EC) = f(Ctc, TMB)$		Model: $_{\log10}(EC) = f(Ctc, BSn, TMB)$	
	n	126	*n*	152
	Const	0.0012 (0.0009)	Const	0.0010 (0.0015)
	Ctc	0.3140 (0.0002)	*Ctc*	0.7086 (≤0.0001)
	TMB	0.2847 (0.3851)	*BSn*	0.2916 (0.9015)
	*Ctc*TMB*	0.1058	*TMB*	0.2869 (0.1524)
			*Ctc*BSn*	0.6879
			*Ctc*TMB*	0.1383
			*BSn*TMB*	0.5190
			*Ctc*BSn*TMB*	0.3294
All flows	Models: $\log_{10}(EC) = f(Ctc, Qxquart)$		Model: $\log_{10}(EC) = f(Ctc, Qquart, BSN)$	
	n	399	*n*	725
	Const	≤0.0001 (≤0.0001)	Const	≤0.0001 (0.0001)
	Ctc	0.0026 (≤0.0001)	*Ctc*	0.0015 (≤0.0001)
	Qquart	≤0.0001 (≤0.0001)	*Qquart*	≤0.0001 (≤0.0001)
	*Ctc*Qquart*	0.0038	*BSn*	0.0879 (≤0.0001)
			*Ctc*Qquart*	0.0007
	Const	≤0.0001 (≤0.0001)	*Ctc*BSn*	0.9753
	Ctc	0.0017 (≤0.0001)	*Qquart*BSn*	0.1484
	Q1quart	≤0.0001 (≤0.0001)	*Ctc*Qquart*BSn*	0.0232
	*Ctc*Q1quart*	0.0222		
	Const	≤0.0001 (≤0.0001)		
	Ctc	≤0.0001 (≤0.0001)		
	Q2quart	0.0009 (0.0015)		
	*Ctc*Q2quart*	0.4074		
	Model: $\log_{10}(EC) = f(Ctc, TMB)$		Model: $\log_{10}(EC) = f(Ctc, TMB, BSn)$	
	n	346	*n*	633
	Const	0.0042 (0.0065)	Const	0.0062 (0.0085)
	Ctc	0.1197 (≤0.0001)	*Ctc*	0.0193 (≤0.0001)
	TMB	≤0.0001 (≤0.0001)	*TMB*	≤0.0001 (0.0001)
	*Ctc*TMB*	0.2099	*BSn*	0.7427 (≤0.0001)
			*Ctc*TMB*	0.5419
			*Ctc*BSn*	0.4452
			*TMB*BSn*	0.3538
			*Ctc*TMB*BSn*	0.0260

Notes: *Ctc* = Catchment Type (**B**, **D**, **F**, **M**, **S**). *BSn* = Bathing Season (Yes or No). *TMB* = Turbidity (C or T), assessed by G McBride from the turbidity records. *Qquart* = flow quartile on sampling day. *Qxquart* = flow quartile on day *x* before sampling. *Ctc*, *BSn*, are fixed factors; *TMB*, *Qquart*, *Qxquart* are random factors. Turbidity and flow not included simultaneously, as they are correlated.

Models for *C. perfringens* spores

	Bathing season only		All times	
Low flows	Model: $\log_{10}(CPS) = f(Ctc)$		Model: $\log_{10}(CPS) = f(Ctc, BSn)$	
	n	176	*n*	241
	Const	0.2851	Const	0.0061 (0.0061)
	Ctc	0.0012	*Ctc*	≤0.0001 (≤0.0001)
			BSn	0.0552 (0.0192)
			*Ctc*BSn*	0.4845
	Model: $\log_{10}(CPS) = f(Ctc, TMB)$		Model: $\log_{10}(CPS) = f(Ctc, BSn, TMB)$	
	n	126	*n*	152
	Const	0.0859 (0.1047)	Const	0.0893 (0.0140)
	Ctc	0.5102 (0.0109)	*Ctc*	0.1047 (0.0261)
	TMB	0.5050 (0.4088)	*BSn*	0.8557 (0.6202)
	*Ctc*TMB*	0.9426	*TMB*	0.4334 (0.8907)
			*Ctc*BSn*	0.0770
			*Ctc*TMB*	0.9921
			*BSn*TMB*	0.1530
			*Ctc*BSn*TMB*	0.9028
All flows	Models: $\log_{10}(CPS) = f(Ctc, Qxquart)$		Model: $\log_{10}(CPS) = f(Ctc, Qquart, BSN)$	
	n	399	*n*	725
	Const	0.2706 (0.2454)	Const	0.1126 (0.1293)
	Ctc	0.0150 (≤0.0001)	*Ctc*	0.0485 (≤0.0001)
	Qquart	≤0.0001 (≤0.0001)	*Qquart*	≤0.0001 (≤0.0001)
	*Ctc*Qquart*	0.0112	*BSn*	0.5654 (0.1093)
			*Ctc*Qquart*	0.0007
	Const	0.2218 (0.1824)	*Ctc*BSn*	0.9087
	Ctc	0.0075 (≤0.0001)	*Qquart*BSn*	0.0832
	Q1quart	≤0.0001 (≤0.0001)	*Ctc*Qquart*BSn*	0.0844
	*Ctc*Q1quart*	0.0112		
	Const	0.1127 (0.1094)		
	Ctc	0.0079(≤0.0001)		
	Q2quart	0.0018 (0.0025)		
	*Ctc*Q2quart*	0.1334		
	Model: $\log_{10}(CPS) = f(Ctc, TMB)$		Model: $\log_{10}(CPS) = f(Ctc, TMB, BSn)$	
	n	346	*n*	633
	Const	0.2810 (0.2790)	Const	0.1697 (0.1613)
	Ctc	0.5144 (≤0.0001)	*Ctc*	0.8550 (≤0.0001)
	TMB	0.0282 (0.0287)	*TMB*	≤0.0001 (≤0.0001)
	*Ctc*TMB*	0.6536	*BSn*	0.7388 (0.0003)
			*Ctc*TMB*	0.0168
			*Ctc*BSn*	0.8648
			*TMB*BSn*	0.3430
			*Ctc*TMB*BSn*	0.7127

Notes: *Ctc* = Catchment Type (**B**, **D**, **F**, **M**, **S**). *BSn* = Bathing Season (Yes or No). *TMB* = Turbidity (C or T), assessed by G McBride from the turbidity records. *Qquart* = flow quartile on sampling day. *Qxquart* = flow quartile on day *x* before sampling. *Ctc*, *BSn*, are fixed factors; *TMB*, *Qquart*, *Qxquart* are random factors. Turbidity and flow not included simultaneously, as they are correlated.

Models for somatic coliphage

	Bathing season only		**All times**	
Low flows	Model: $\log_{10}(SP) = f(Ctc)$		Model: $\log_{10}(SP) = f(Ctc, BSn)$	
	n	176	*n*	241
	Const	≤0.0001	Const	≤0.0001 (≤0.0001)
	Ctc	≤0.0001	*Ctc*	≤0.0001 (≤0.0001)
			BSn	0.7742 (0.8885)
			*Ctc*BSn*	0.1151
	Model: $\log_{10}(SP) = f(Ctc, TMB)$		Model: $\log_{10}(SP) = f(Ctc, BSn, TMB)$	
	n	126	*n*	152
	Const	0.0025 (0.0094)	Const	0.0018 (0.0086)
	Ctc	0.3857 (≤0.0001)	*Ctc*	0.4880 (≤0.0001)
	TMB	0.4920 (0.0824)	*BSn*	0.5161 (0.2205)
	*Ctc*TMB*	0.1057	*TMB*	0.5457 (0.0577)
			*Ctc*BSn*	0.8285
			*Ctc*TMB*	0.0577
			*BSn*TMB*	0.3273
			*Ctc*BSn*TMB*	0.0759
All flows	Models: $\log_{10}(SP) = f(Ctc, Qxquart)$		Model: $\log_{10}(SP) = f(Ctc, Qquart, BSN)$	
	n	399	*n*	725
	Const	0.0001 (0.0003)	Const	≤0.0001 (0.0003)
	Ctc	≤0.0001 (≤0.0001)	*Ctc*	0.0001 (≤0.0001)
	Qquart	0.0006 (≤0.0001)	*Qquart*	0.0002 (≤0.0001)
	*Ctc*Qquart*	0.0067	*BSn*	0.8698 (0.1308)
			*Ctc*Qquart*	≤0.0001
	Const	≤0.0001 (0.0001)	*Ctc*BSn*	0.8412
	Ctc	0.0003 (≤0.0001)	*Qquart*BSn*	0.4114
	Q1quart	0.0170 (0.0011)	*Ctc*Qquart*BSn*	0.0085
	*Ctc*Q1quart*	0.0029		
	Const	≤0.0001 (0.0001)		
	Ctc	≤0.0001(≤0.0001)		
	Q2quart	0.1774 (0.0032)		
	*Ctc*Q2quart*	0.1279		
	Model: $\log_{10}(SP) = f(Ctc, TMB)$		Model: $\log_{10}(SP) = f(Ctc, TMB, BSn)$	
	n	346	*n*	633
	Const	0.0042 (0.0083)	Const	0.0083 (0.0120)
	Ctc	0.0274 (≤0.0001)	*Ctc*	0.0029 (≤0.0001)
	TMB	0.0274 (0.0008)	*TMB*	≤0.0001 (≤0.0001)
	*Ctc*TMB*	0.4034	*BSn*	0.8209 (0.4540)
			*Ctc*TMB*	0.7754
			*Ctc*BSn*	0.3922
			*TMB*BSn*	0.2632
			*Ctc*TMB*BSn*	0.0256

Notes: *Ctc* = Catchment Type (**B**, **D**, **F**, **M**, **S**). *BSn* = Bathing Season (Yes or No). *TMB* = Turbidity (C or T), assessed by G McBride from the turbidity records. *Qquart* = flow quartile on sampling day. *Qxquart* = flow quartile on day *x* before sampling. *Ctc*, *BSn*, are fixed factors; *TMB*, *Qquart*, *Qxquart* are random factors. Turbidity and flow not included simultaneously, as they are correlated.

Models for FRNA phage

	Bathing season only		**All times**	
Low flows	Model: $\log_{10}(FRNA) = f(Ctc)$		Model: $\log_{10}(FRNA) = f(Ctc, BSn)$	
	n	176	*n*	241
	Const	0.1475	Const	0.1679 (0.1724)
	Ctc	0.0044	*Ctc*	0.0060 (0.0022)
			BSn	0.8147 (0.6230)
			*Ctc*BSn*	0.0629
	Model: $\log_{10}(FRNA) = f(Ctc, TMB)$		Model: $\log_{10}(FRNA) = f(Ctc, BSn, TMB)$	
	n	126	*n*	152
	Const	0.4970 (0.5561)	Const	0.1396 (0.5085)
	Ctc	0.3732 (0.0003)	*Ctc*	0.0233 (≤0.0001)
	TMB	0.0475 (0.0131)	*BSn*	0.1266 (0.1334)
	*Ctc*TMB*	0.9295	*TMB*	0.7007 (0.0553)
			*Ctc*BSn*	0.0340
			*Ctc*TMB*	0.8905
			*BSn*TMB*	0.5205
			*Ctc*BSn*TMB*	0.9316
All flows	Models: $\log_{10}(FRNA) = f(Ctc, Qxquart)$		Model: $\log_{10}(FRNA) = f(Ctc, Qquart, BSN)$	
	n	399	*n*	725
	Const	0.2469 (0.2805)	Const	0.0595 (0.1861)
	Ctc	0.0010 (≤0.0001)	*Ctc*	0.0004 (≤0.0001)
	Qquart	≤0.0001 (≤0.0001)	*Qquart*	≤0.0001 (≤0.0001)
	*Ctc*Qquart*	0.0175	*BSn*	0.2778 (0.8767)
			*Ctc*Qquart*	0.0003
	Const	0.2348 (0.2452)	*Ctc*BSn*	0.7921
	Ctc	0.0014 (≤0.0001)	*Qquart*BSn*	0.0167
	Q1quart	≤0.0001 (≤0.0001)	*Ctc*Qquart*BSn*	0.0223
	*Ctc*Q1quart*	0.0104		
	Const	0.2329 (0.2499)		
	Ctc	0.0002 (≤0.0001)		
	Q2quart	≤0.0001 (≤0.0001)		
	*Ctc*Q2quart*	0.1657		
	Model: $\log_{10}(FRNA) = f(Ctc, TMB)$		Model: $\log_{10}(FRNA) = f(Ctc, TMB, BSn)$	
	n	346	*n*	633
	Const	0.1818 (0.1298)	Const	0.2012 (0.1736)
	Ctc	0.0835 (≤0.0001)	*Ctc*	0.0481 (≤0.0001)
	TMB	0.0063 (0.0378)	*TMB*	≤0.0001 (≤0.0001)
	*Ctc*TMB*	0.1701	*BSn*	0.7337 (0.0705)
			*Ctc*TMB*	0.0114
			*Ctc*BSn*	0.5915
			*TMB*BSn*	0.1206
			*Ctc*TMB*BSn*	0.7071

Notes: *Ctc* = Catchment Type (**B**, **D**, **F**, **M**, **S**). *BSn* = Bathing Season (Yes or No). *TMB* = Turbidity (C or T), assessed by G McBride from the turbidity records. *Qquart* = flow quartile on sampling day. *Qxquart* = flow quartile on day *x* before sampling. *Ctc*, *BSn*, are fixed factors; *TMB*, *Qquart*, *Qxquart* are random factors. Turbidity and flow not included simultaneously, as they are correlated.

Models for *Campylobacter* (with turbidity)

	Bathing season only		**All times**	
Low flows	Model: $\log_{10}(Camp) = f(Ctc)$		Model: $\log_{10}(Camp) = f(Ctc, BSn)$	
	n	176	*n*	240
	Const	≤0.0001	Const	≤0.0001 (≤0.0001)
	Ctc	0.5688	*Ctc*	0.2666 (0.3538)
			BSn	0.2113 (0.2446)
			*Ctc*BSn*	0.6327
	Model: $\log_{10}(Camp) = f(Ctc, TMB)$		Model: $\log_{10}(Camp) = f(Ctc, BSn, TMB)$	
	n	126	*n*	151
	Const	0.0411 (0.0486)	Const	0.0563 (0.2659)
	Ctc	0.7995 (0.4440)	*Ctc*	0.8864 (0.1779)
	TMB	0.8132 (0.7840)	*BSn*	0.1633 (0.0245)
	*Ctc*TMB*	0.2695	*TMB*	0.8669 (0.3747)
			*Ctc*BSn*	0.1878
			*Ctc*TMB*	0.3264
			*BSn*TMB*	0.9610
			*Ctc*BSn*TMB*	0.8015
All flows	Models: $\log_{10}(Camp) = f(Ctc, Qxquart)$		Model: $\log_{10}(Camp) = f(Ctc, Qquart, BSN)$	
	n	399	*n*	726
	Const	0.1132 (0.2198)	Const	0.6804 (0.8199)
	Ctc	0.0119 (≤0.0001)	*Ctc*	0.0153 (≤0.0001)
	Qquart	0.0167 (0.0001)	*Qquart*	0.0135 (≤0.0001)
	*Ctc*Qquart*	0.1917	*BSn*	0.1950 (0.0023)
			*Ctc*Qquart*	0.0402
	Const	0.0613 (0.1971)	*Ctc*BSn*	0.9103
	Ctc	0.0087 (0.0001)	*Qquart*BSn*	0.3929
	Q1quart	0.1204 (0.0005)	*Ctc*Qquart*BSn*	0.2513
	*Ctc*Q1quart*	0.5859		
	Const	0.0676 (0.1922)		
	Ctc	0.0197(0.0001)		
	Q2quart	0.0941 (0.0006)		
	*Ctc*Q2quart*	0.2544		
	Model: $\log_{10}(Camp) = f(Ctc, TMB)$		Model: $\log_{10}(Camp) = f(Ctc, TMB, BSn)$	
	n	346	*n*	633
	Const	0.1760 (0.2024)	Const	0.7836 (0.8177)
	Ctc	0.6274 (0.0003)	*Ctc*	0.1117 (≤0.0001)
	TMB	0.4100 (0.3408)	*TMB*	0.0011 (≤0.0001)
	*Ctc*TMB*	0.3896	*BSn*	0.7675 (0.0016)
			*Ctc*TMB*	0.8885
			*Ctc*BSn*	0.4099
			*TMB*BSn*	0.0722
			*Ctc*TMB*BSn*	0.8896

Notes: *Ctc* = Catchment Type (**B**, **D**, **F**, **M**, **S**). *BSn* = Bathing Season (Yes or No). *TMB* = Turbidity (C or T), assessed by G McBride from the turbidity records. *Qquart* = flow quartile on sampling day. *Qxquart* = flow quartile on day *x* before sampling. *Ctc*, *BSn*, are fixed factors; *TMB*, *Qquart*, *Qxquart* are random factors. Turbidity and flow not included simultaneously, as they are correlated.

Models for *Campylobacter* (with *E. coli*)

	Bathing season only		**All times**	
Low flows	Model: $\log_{10}(Camp) = f(Ctc)$		Model: $\log_{10}(Camp) = f(Ctc, BSn)$	
	n	176	*n*	240
	Const	≤0.0001	Const	≤0.0001 (≤0.0001)
	Ctc	0.5688	*Ctc*	0.2666 (0.3538)
			BSn	0.2113 (0.2446)
			*Ctc*BSn*	0.6327
	Model: $\log_{10}(Camp) = f[Ctc, \log_{10}(EC)]$		Model: $\log_{10}(Camp) = f[Ctc, BSn, \log_{10}(EC)]$	
	n	126	*n*	240
	Const	0.3330 (0.3669)	Const	0.5054 (0.5017)
	Ctc	0.5397 (0.8087)	*Ctc*	0.5700 (0.2720)
	$\text{Log}_{10}(EC)$	0.0125 (0.0048)	*BSn*	0.6163 (0.0471)
	*Ctc*EC*	0.6506	$\text{Log}_{10}(EC)$	≤0.0001 (≤0.0001)
			*Ctc*BSn*	0.5007
			*Ctc** $\text{Log}_{10}(EC)$	0.0402
			*BSn** $\text{Log}_{10}(EC)$	0.0344
			*Ctc*BSn** $\text{Log}_{10}(EC)$	0.4837
All flows	Models: $\log_{10}(Camp) = f(Ctc, Qxquart)$		Model: $\log_{10}(Camp) = f(Ctc, Qquart, BSN)$	
	n	399	*n*	726
	Const	0.1132 (0.2198)	Const	0.6804 (0.8199)
	Ctc	0.0119 (≤0.0001)	*Ctc*	0.0153 (≤0.0001)
	Qquart	0.0167 (0.0001)	*Qquart*	0.0135 (≤0.0001)
	*Ctc*Qquart*	0.1917	*BSn*	0.1950 (0.0023)
			*Ctc*Qquart*	0.0402
	Const	0.0613 (0.1971)	*Ctc*BSn*	0.9103
	Ctc	0.0087 (0.0001)	*Qquart*BSn*	0.3929
	Q1quart	0.1204 (0.0005)	*Ctc*Qquart*BSn*	0.2513
	*Ctc*Q1quart*	0.5859		
	Const	0.0676 (0.1922)		
	Ctc	0.0197(0.0001)		
	Q2quart	0.0941 (0.0006)		
	*Ctc*Q2quart*	0.2544		
	Model: $\log_{10}(Camp) = f[Ctc, \log_{10}(EC)]$		Model: $\log_{10}(Camp) = f[Ctc, \log_{10}(EC), BSn]$	
	n	399	*n*	724
	Const	0.6909 (0.6927)	Const	0.9530 (0.9534)
	Ctc	0.7235 (0.0531)	*Ctc*	0.6670 (0.0008)
	$\text{Log}_{10}(EC)$	≤0.0001 (≤0.0001)	$\text{Log}_{10}(EC)$	≤0.0001 (≤0.0001)
	*Ctc** $\text{Log}_{10}(EC)$	0.0183	*BSn*	0.7835 (≤0.0001)
			*Ctc** $\text{Log}_{10}(EC)$	≤0.0001
			*Ctc*BSn*	0.5018
			$\text{Log}_{10}(EC)$**BSn*	0.0011
			*Ctc** $\text{Log}_{10}(EC)$**BSn*	0.8823

Notes: *Ctc* = Catchment Type (**B**, **D**, **F**, **M**, **S**). *BSn* = Bathing Season (Yes or No). *TMB* = Turbidity (C or T), assessed by G McBride from the turbidity records. *Qquart* = flow quartile on sampling day. *Qxquart* = flow quartile on day *x* before sampling. *Ctc*, *BSn*, are fixed factors; *TMB*, *Qquart*, *Qxquart* are random factors. Turbidity and flow not included simultaneously, as they are correlated.

The results for the *Campylobacter* model in the south east box of the previous table show that *E. coli* features importantly in a model for *Campylobacter*. The converse is also true, as shown in the results below.

Model: $\log_{10}(EC) = f[Ctc, \log_{10}(Camp), BSn]$	
n	724
Const	0.0917 (00965)
Ctc	0.0775 (≤0.0001)
$\text{Log}_{10}(EC)$	≤0.0001 (≤0.0001)
BSn	0.1596 (≤0.0001)
*Ctc** $\text{Log}_{10}(EC)$	0.0004
*Ctc***BSn*	0.0461
$\text{Log}_{10}(EC)$**BSn*	0.1850
*Ctc** $\text{Log}_{10}(EC)$**BSn*	0.9970

Models for Adenovirus and enterovirus

These variables are dichotomous (present/absent) and so lend themselves to logistic models, of the form logit(virus presence/absence) = a linear function of catchment type, bathing season and turbidity, i.e., the case considered in the southeast corner of the preceding tables (except the last one).[55]

Results (*p*-values for 633 values, i.e., all those with accompanying turbidity measurements) are shown below.

Term	Adenovirus	Enterovirus
Const	0.0060	0.0003
Ctc	0.0002	0.6408
BSn	0.6027	0.5515
TMB	0.2629	0.9023

Note the small value for catchment type in the adenovirus column. This is saying that adenovirus does tend to vary strongly between catchments, even when accounting for turbidity and bathing season. That is not true for enterovirus.

The cyst data are too sparse to lend themselves to this modelling,[56] as evidenced by the iterative algorithm used in logistic modelling often failing to converge.

[55] The logit of a proportion of positive results (denoted by *p*) is $\ln[p/(1-p)]$. Making this equal to a linear combination of factors gives rise to a logistic curve, rising from 0 to 1.

[56] There being so few positive cyst results it only makes sense to model them as dichotomous variables.

A.3.7 Modelling health risks

Microorganisms modelled
Six pathogens were assayed in the FMRP (*Campylobacter*, adenoviruses, enteroviruses, *Salmonella*, *Giardia* cysts and *Cryptosporidium* oocysts). However, the data collected demonstrated elevated levels of the first three only. We therefore confine the modelling to these three.[57] This modelling is done for all data, for all data in the bathing season (1 November – 30 March), and for data in each of the five catchment types.

Some general discussion of approaches to the modelling health risks from exposure to pathogenic microorganisms is warranted. This also includes a discussion of the populations at risk and of the special issues posed by the use of MPN results (for *Campylobacter*) and presence/absence results (for viruses).

Dose-response models for infection
As agreed by the Risk Assessment Working Group, we confine health effects modelling to *infection*, rather than *illness*. This means that we can use the dose-response data available in the literature for rates of infection, but it is then more difficult to use epidemiological data (on outbreaks, and in cohort studies of various sorts) because these are generally confined to measures of illness.

A considerable amount of dose-response data for infection is reported in the literature from the results of clinical trials (summarised in Haas *et al*. 1999, and in Fewtrell & Bartram 2001),[58] e.g., for:

- campylobacteriosis — Black *et al*. (1988)
- adenovirus 4 infection — Couch *et al*. (1966a[59] & b), Couch *et al*. 1969
- echovirus 12 infection[60] — Akin (1981, cited in Haas 1983 and in Haas *et al*. 1999); Schiff *et al*. (1984a&b)
- salmonellosis — McCullough & Eisele (1951a&b)
- cryptosporidiosis — Dupont *et al.* (1995); Chappell *et al.* (1996); Moss *et al.* (1998); Okhuysen *et al*. (1998, 1999); Messner *et al*. (2001)
- giardiasis — Rendtorff (1954a&b), Rendtorff & Holt (1954a&b)
- cholera — Hornick *et al*. (1971)
- rotavirus infection — Ward *et al*. (1986)
- coxsackie (B4 and A21) infection — Couch *et al*. (1965), Suptel (1963)

Fortunately, a great deal of work has been done to see how best to fit these data into dose-response models (as summarised in the text by Haas *et al*. 1999); that is, how do we relate the probability of infection (denoted as P_{inf}) to the mean dose (denoted as N)?[61] These models fall into one of two types. The first, and simpler, is the "exponential model"

$$P_{\text{inf}} = 1 - e^{-rN} \qquad (1)$$

[57] Note that were cysts and oocysts data to be analysed, special care would be needed in handling non-detects in samples (Parkhurst & Stern 1998).
[58] Populations used typically comprised healthy adults, e.g., prisoners and military personnel (McCullough & Eisele 1947, Rendtorff 1954, Knight 1964) and university staff and students (DuPont 1995). See also excellent summaries of gastro-intestinal dose-response studies by Teunis *et al*. (1996).
[59] Title incorrect in Haas *et al.* (1999).
[60] This may be taken as a typical enterovirus, as did DRG (2001).
[61] Some studies (e.g., the campylobacteriosis study by Black *et al.* 1988) include measures of illness as well as of infection, but the modelling of illness-given-infection is challenging (e.g., Teunis *et al*. 1999). With the use of literature data from volunteer experiments, examples can be found for three possible alternatives: an increase in the probability of illness with increasing dose (salmonellosis), a decrease with higher doses (campylobacteriosis), and a probability of illness (given infection) independent of the ingested dose (cryptosporidiosis). These alternatives may reflect different modes of interactions between pathogens and hosts (Teunis *et al*. 1996).

where each micro-organism has the same fixed probability (r) of surviving and reaching a host site at which infection may be the result (so $0 < r < 1$). Under this model the ID_{50} is $N_{50} = -\ell n(0.5)/r = 0.693/r$.[62] Protozoa and many viruses tend to follow this model. That is, differential susceptibility in the (healthy) population when challenged by potentially harmful viruses and cysts tends not to be strong.[63]

The three key assumptions made to develop the exponential model are:

- microorganisms are distributed randomly within a water volume and so follow the Poisson distribution;
- for infection to occur *at least one* pathogen must survive within the host (so it is a "single hit" model), characterised as "$k_{min} = 1$" (see Haas *et al.* 1999, p. 263);[64]
- the probability of infection per ingested or inhaled organism (r) is constant.[65]

Note that k_{min} is *not* the "minimum infectious dose"—infection is described by the joint action of k_{min} and r. The former term ($k_{min} = 1$) refers to an ingested or inhaled organism reaching a site where it, and it alone, can cause infection. However not all ingested or inhaled organisms "… reach a target site where they can initiate a response, because they may not be viable, they may be inactivated by decay, or their activity may be impaired due to host defences." (Teunis *et al*. 1995). The latter term (r) accounts for the success rate of an organism actually reaching a site within the host where infection could occur.[66] The range of values of r reported in the literature is given in Table A3.7.1.

The second approach is the "beta-Poisson model". This model abandons the third assumption made in developing the single-hit exponential model. It recognises that the probability of infection per ingested or inhaled organism of an ingested particle may not be constant in the population at large and instead makes r follow a beta distribution, in which case the model no longer contains r but instead contains the two parameters (α and β) of the beta distribution that describe the variation in r,[67] i.e.,

$$P_{\text{inf}} = 1 - \left(1 + \frac{N}{\beta}\right)^{-\alpha} \qquad (2)$$

Bacteria tend to conform to this model.[68] In other words, there is substantial differential susceptibility among the (healthy) population at large when challenged with potentially harmful bacteria. Note that equation 2 is an approximation to the correct result[69] and so may not conform well to a single-hit model at low doses (Teunis & Havelaar 2000)—this is not expected to be an issue in the current application to *Campylobacter*.[70] Note too that the beta-Poisson model can be reformulated (e.g., Regli *et al.* 1991, Haas *et al*. 1999) as

---

[62] This is the dose require to cause infection in half of an exposed population, obtained by setting $P_{inf} = ½$ in the exponential model equation. Confusingly, the ID_{50} is sometimes called the median infectious dose (Haas *et al*. 1999) and sometimes the "minimal infective dose" (Haas *et al.* 1999, p. 263) or "minimum infectious dose" (Ward & Akin 1984), whereas others (e.g., Schiff *et al*. 1984a&b) define "minimal infectious dose" as that required to infect just 1% of an exposed population.

[63] Note that clinical trial data is available for Norwalk virus infection (Graham *et al*. 1994), an important agent of gastro-intestinal illness in environmental waters, but difficulties in its enumeration prevent a dose-response relation being identified.

[64] If more than one are required then $k_{min} > 1$ and the model is no longer "single-hit".

[65] Under a single hit model r is both the probability of a microorganism surviving and reaching a host site at which infection may be the result, or, equivalently, it is the probability of infection per ingested or inhaled organism.

[66] Multi-hit models (i.e., where $k_{min} > 1$) are seldom used in microbial risk assessment (Haas *et al*. 1999, p. 271).

[67] The beta distribution has just two dimensionless parameters (α and β). According to the values given to those parameters it can assume many shapes: left-skewed, right-skewed, U-shaped, humped, or monotonic increasing or monotonic decreasing (an excellent pictorial summary is given in Lee 1989).

[68] Strictly, in the beta-Poisson model one should have $\beta >> \alpha$ (the development of the beta-Poisson equation requires this condition to be met, and it also guarantees that the beta distribution function is monotonic decreasing as $r \to 1$). In practice this condition is not always met (Teunis *et al*. 1995), e.g., for Rotavirus and Poliovirus III—see Table A3.7.1.

[69] This is a confluent hypergeometric function, see Haas *et al*. 1999, page 267.

[70] It would be were the risk modelling to include sensitivity to the dose-response curve parameters: Teunis & Havelaar 2000 show that while the exact and approximate curves agree almost exactly, their confidence intervals at low doses do not. Indeed the exact curve's interval is much narrower at low doses than is the beta-Poisson curve. As stated in footnote 27, this issue is to be studied further.

$$P_{\text{inf}} = 1 - \left[1 + \frac{N}{N_{50}} \left(2^{\frac{1}{\alpha}} - 1 \right) \right]^{-\alpha} \qquad (3)$$

where N_{50} is the ID_{50}. This formulation is obtained by setting $P_{\text{inf}} = ½$ in the beta-Poisson equation—and so $\beta = N_{50}/(2^{1/\alpha}-1)$.[71] Unfortunately, α appears not to have an obvious physical interpretation in the (approximate) beta-Poisson model. What can be said of it is that it is a shape parameter governing the steepness of the dose-response curve; the larger its value the steeper the curve; whatever its value the curve always passes through the point ($N = N_{50}$, $P_{\text{inf}} = ½$).

Depending on the model used we need to calibrate either one parameter (r in the exponential model) or two parameters (α and β—or, equivalently, α and N_{50}) in the beta-Poisson model. The range of these parameters reported in the literature is given in Table A3.7.1.

Note that many of the parameter values reported in this table are based on the same dose-response data (e.g., the only available data for *Campylobacter jejuni* is that given by Black *et al.* 1988), and this is often (but not always) reflected in similar parameter estimates in the various papers cited.

The parameter values shown in this table were obtained by fitting exponential and beta-Poisson models to experimental dose-response data using human subjects. It is to be noted that for the most part these data were obtained for healthy adults (who do comprise the large majority of recreational water users), but the very young, the very old and immune-compromised populations are not well represented by these data.[72]

It should also be noted that dose-response studies that supply the data for these parameter estimates are generally confined to particular strains of the microorganism in question, but there may be substantial between-strain infectivities.[73] This needs to realised when interpreting the only available *Campylobacter jejuni* infectivity study, that of Black *et al*. (1988). This study included only one strain (A3249, Penner serotype 27) with a sufficient range of doses to establish the estimates of α and N_{50}.[74] Furthermore, earlier studies on a strain of *Cryptosporidium parvum* (Dupont *et al*. 1995) used a strain (IOWA) that was less infective than another (TAMU) strain used in a later study.[75] For this reason the value of r given at the end of Table A3.7.1 (for ref. 11) is higher than that in the middle (for ref. 7).[76] And as another example, the value of r reported in the table for Echovirus 12 by Haas *et al*. (1999, using data from Akin 1981) corresponds to an ID_{50} of 54 particles, whereas Schiff *et al*. (1984a) reported an ID_{50} of 919 particles for the strain they isolated in a different study (these data were also used by Rose & Sobsey 1993 and Teunis *et al* 1996 to calibrate dose-response curves—see also footnote 79). Adopting a precautionary approach, we have adopted the dose-response relation reported by Haas *et al*. (1999).

Table A3.7.1 Calibrated dose-response parameters, in date order

[71] There is an error in this equation in Haas *et al*. (1999)—their eq. 7-20.

[72] Theoretical (Esrey *et al*. 1985) and epidemiological (Ani *et al*. 1988; Figueroa *et al.* 1989) studies concerning young children are an exception. Dose-response studies on infants and newborns have been carried out for Poliovirus (summarised in Teunis *et al*. 1996).

[73] As reported for *in vivo* and *in vitro* laboratory studies between strains of *C. jejuni* (Stewart-Tull *et al*. 1999; Bacon *et al*. 1999).

[74] The 4 Complex Penner serotype is more commonly associated with disease in New Zealand, although Type 27 also figures in disease analysis (C. Nicholl ESR, pers. comm.). More recent human volunteer studies reported from the USA Naval Medical Research Centre, Forest Glen, MD have used the other strain reported by Black *et al*. (1988), i.e., Strain 81-176 Penner serotyoe 23/36, but also at doses too high to adduce a dose-response relationship (Rollins *et al*. 1999; Tribble *et al*. 1996).

[75] These studies have all been carried out at the University of Texas Medical School at Houston. The IOWA strain was collected from naturally infected calves during a diarrhoeal episode and passaged in the laboratory setting for a number of years. The TAMU strain was collected from a veterinary student who became infected while participating in a necropsy on an infected foal. This isolate had been passaged twice prior to the volunteer dose-response studies (Messner *et al*. 2001). These are all "Type 2" strains—originating from animals.

[76] Further studies at Houston have now been carried out on a Type 1 strain, with results similar to the pooled ID_{50} estimates from the Type 2 strains (G. Medema, KIVA, The Netherlands, pers. comm.)—results yet to be published.

Ref.	Stated Model†	Microorganism	Parameters‡		
1	$P_i = 1 - \exp(-rN)$ $P_i = 1 - [1+(\mu V/N_{50})(2^{1/\alpha}-1)]^{-\alpha}$	Echovirus 12	$\alpha = 0.374$	$\beta = 186.69$	
		Rotavirus	$\alpha = 0.26$	$\beta = 0.42$	
		Poliovirus I	$\alpha = 0.1097$	$\beta = 1524$	
		Poliovirus III	$\alpha = 0.409$	$\beta = 0.788$	
		Giardia lamblia	$r = 0.02$		
		Poliovirus I	$r = 0.009102$		
2	$P_i = 1 - \exp(-rN)$	*Giardia cysts*	r = "–0.01982" (minus sign is an error)		
3	$p = 1 - \exp(-rN)$ $p = 1 - [1+N/\beta]^{-\alpha}$	*Campylobacter*	$\alpha = 0.039$	$\beta = 55$	
		Salmonella	$\alpha = 0.33$	$\beta = 139.9$	
		Salmonella typhi	$\alpha = 0.21$	$\beta = 5,531$	
		Poliovirus 1	$\alpha = 15$	$\beta = 1,000$	
		Poliovirus 3	$\alpha = 0.5$	$\beta = 1.14$	
		Echovirus 12	$\alpha = 1.3$	$\beta = 75$	
		Rotavirus	$\alpha = 0.232$	$\beta = 0.247$	
		Giardia lamblia	$r = 0.0199$		
4	$P_i = 1 - (1 + N/\beta)^{-\alpha}$ $P_i = 1 - \exp(-rN)$	Echovirus 12	$\alpha = 0.374$	$\beta = 186.69$	
		Rotavirus	$\alpha = 0.26$	$\beta = 0.42$	
		Poliovirus I	$\alpha = 0.1097$	$\beta = 1524$	
		Poliovirus III	$\alpha = 0.409$	$\beta = 0.788$	
		Poliovirus I	$r = 0.009102$		
5	$P_{inf} = 1 - (1 + N/\beta)^{-\alpha}$	*Campylobacter jejuni*	$\alpha = 0.145$	$\beta = 7.59$	(so $N_{50} = 896.7$)
6	$P_{inf} = 1 - (1 + N/\beta)^{-\alpha}$	Echovirus 12	$\alpha = 0.401$	$\beta = 227.2$	(so $N_{50} = 1052.5$)
		Campylobacter jejuni	$\alpha = 0.145$	$\beta = 7.589$	(so $N_{50} = 896.7$)
7	$P_i = 1 - (1 + N/\beta)^{-\alpha}$	Rotavirus	$\alpha = 0.26$	$\beta = 0.42$	
8	$p = 1 - \exp(N/k)$	*Cryptosporidium parvum*	$k = 238.6$	(so $r = 0.0042$)	
9	$P^*_{inf} = 1 - e^{-r \times Dose}$	*Giardia*	Not given		
10	$P_i = 1 - \exp(-rN)$	Adenovirus	$r = 0.4172$		
11	$P_I(d) = 1 - \exp(-d/k)$ $P_I(d) = 1 - [1+(d/N_{50})(2^{1/\alpha}-1)]^{-\alpha}$ NB: their eq. 7-20 is in error—"$N_{50} = \beta/(2^{1/\alpha}-1)$" should be "$\beta = N_{50}/(2^{1/\alpha}-1)$". That equation is derived by setting $P_I(d) = ½$ in the overall beta-Poisson equation (eq. 7-19), i.e., $P_I(d) = 1 - [1+(d/\beta)]^{-\alpha}$. Note too that $k = 1/r$.	Rotavirus	$\alpha = 0.2531$	$N_{50} = 6.17$	(so $\beta = 0.4265$)
		Salmonella (nontyphoid)	$\alpha = 0.3126$	$N_{50} = 23,600$	(so $\beta = 2,884$)
		Salmonella typhi	$\alpha = 0.1086$	$N_{50} = 3.6 \times 10^6$	(so $\beta = 6,097$)
		Shigella	$\alpha = 0.2100$	$N_{50} = 1,120$	(so $\beta = 42.86$)
		E. coli (pathogenic strains)	$\alpha = 0.1778$	$N_{50} = 8.6 \times 10^7$	(so $\beta = 1.8 \times 10^6$)
		Campylobacter jejuni	$\alpha = 0.145$	$N_{50} = 896$	(so $\beta = 7.58$)
		Vibrio cholera	$\alpha = 0.25$	$N_{50} = 243$	(so $\beta = 16.2$)
		Endamoeba coli	$\alpha = 0.1008$	$N_{50} = 341$	(so $\beta = 0.3522$)
		Poliovirus I (minor)	$k = 109.87$	(so $r = 0.0091$)	
		Hepatitis A virus	$k = 1.8229$	(so $r = 0.5486$)	
		Adenovirus 4	$k = 2.397$	(so $r = 0.4172$)	
		Echovirus 12	$k = 78.3$	(so $r = 0.0128$)	
		Coxsackie virus	$k = 69.1$	(so $r = 0.0145$)	
		Cryptosporidium parvum	$k = 238$	(so $r = 0.0042$)	
		Giardia lamblia	$k = 50.23$¶	(so $r = 0.0199$)	
12	$ID_{50} = 9$ (TAMU isolate)	*Cryptosporidium parvum*	$k = 12.99$	(so $r = 0.077$)	

† The model as presented by these authors is shown, to aid comparison with the different nomenclatures used.

‡ For clarity of presentation where both exponential and beta-Poisson models have been used we place the beta-Poisson model parameters first, followed by the exponential model's parameters.

¶ Stated as "50, 23", but should presumably be $k = 50.23$, as in Rose & Gerba (1991)—reference 3.

References:

1 Regli *et al*. (1991)
2 Rose *et al*. (1991)
3 Rose & Gerba (1991)
4 Rose & Sobsey (1993)
5 Medema *et al*. (1996)
6 Teunis *et al*. (1996)
7 Gerba *et al*. (1996)
8 Haas *et al*. (1996)
9 Teunis *et al.* (1997)
10 Crabtree *et al*. (1997)
11 Haas *et al*. (1999)
12 Okhuysen *et al*. (1999)

Note:

These calibrations are based on available data that have used particular pathogen strains processed in particular ways. Where more than one strain of an organism has been studied in clinical trials, a wide range of infectivities can be discovered (e.g., for cryptosporidiosis—Teunis *et al* 2002a&b). *Therefore it must be recognised that these calibrations can carry a substantial degree of uncertainty.*

Figures A3.7.1 and A3.7.2 show dose-response curves to be used later in this analysis. It should be noted that there is little published data available to substantiate some the parameters given for the virus relations given on the latter figure. In particular, the adenovirus curve should be treated with caution (P. Teunis, RIVM, The Netherlands, pers. comm.)—it is based on a very limited number of subjects in clinical trials (Couch *et al.* 1966a).

Figure A3.7.1 shows the initial steepness of the *Campylobacter* curve, up to 100 particles, even though the ID_{50} is N_{50} = 896. Figure A3.7.2 indicates the degree of infectivity of adenoviruses (once two particles are received one has passed the ID_{50}, based on the data used to develop its dose-response relationship). Hepatitis A has a similar infectivity curve.

Monte Carlo modelling of health risk

Modelling health risk is predicated on having a measure of microorganism concentration in the recreational water, a volume of water ingested or inhaled, and a dose-response model. In that case we can make a calculation of the probability of infection, and maybe go on from that to an assessment of probability of illness. This is the approach taken here, as depicted in Figure A3.7.3.

In doing so we note that, technically, there is no such thing as an infectious dose (Gale 1999). It is rather the case that a given concentration of microorganisms has a probability, spread over a population, of causing infection. So, for example, N_{50} is the dose when given to each and every member of a population infects half of that population. This range of probabilities arises because, when considering all these individuals, we need to account for the varying volumes of water they ingest or inhale, and the varying concentrations of microorganisms in that water. It follows that basing risk calculations on average values[77] may underestimate risk, depending on the manner in which average concentrations are determined. This is because for a given average there may be many somewhat lower values and a few much higher values (Gale 1998, Gale & Stansfield 2000); microorganisms are often clustered to some degree within a water body, and this needs to be accounted for in performing risk assessments (because these few high values will contribute disproportionately to the risk, Gale 1996). In the absence of formal analytical methods to account for this feature we use repetitive "brute-force" computational procedures to do the necessary calculations, i.e., Monte Carlo methods.

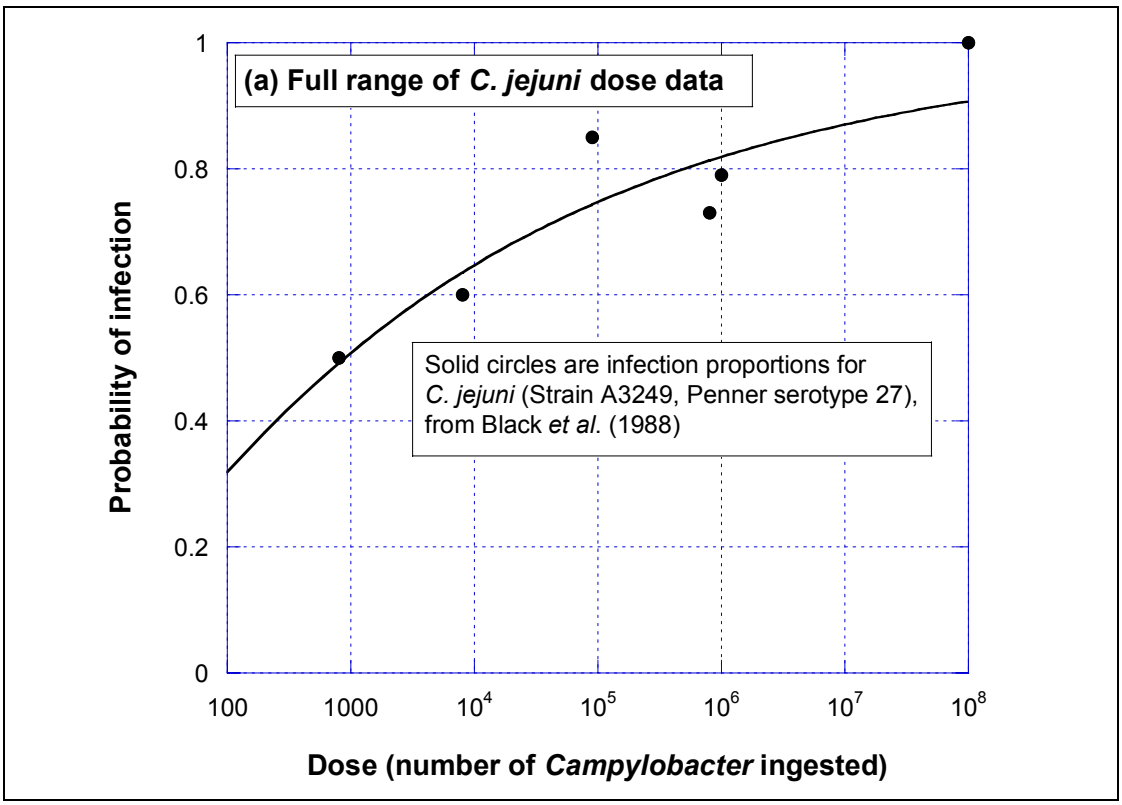
Furthermore, we note that the above-noted clumping of microorganisms implies that the usual "random" (i.e., Poisson) distributions applied to bacteria may not be appropriate and one may need to a more dispersed distribution such as the lognormal or negative binomial (Gale et al. 1997), or an "added zeroes" distribution. Accordingly, in our health risk modelling we are using a Monte-Carlo probabilistic approach. In this we take a range of probabilities for three key variables. We use the betaPERT distribution for the first two probabilities, and a variety of others for the microorganism densities.[78] In detail, we take

- *duration of swimming* as a betaPERT distribution with minimum and maximum durations ¼ and 2 hours (e.g., for wind surfers), with mode (i.e., most likely value) = ½ hour.
- *volume ingested/inhaled per hour* as a betaPERT distribution with minimum and maximum volumes = 10 and 100 mL, with mode = 50 mL.[79]
- *concentration of microorganisms* as from statistical distributions that best fit the data collected in the Programme (i.e., geometric, negative binomial and "Poisson with added zeroes" distributions).

[77] As did Savill *et al*. (2000, 2001).

[78] The betaPERT distribution is related to the beta distribution, but allows the user to specify minimum, maximum and most likely (i.e., mode) values. (It is available as the "PERT" distribution in @RISK, Pallisade Corp. 2000.)

[79] The upper limits refer to swimmers; other recreational users (water skiers, wind surfers) tend to ingest or inhale less water (G. Lewis, University of Auckland, pers. comm.). Note too that Schernewski & Jülich (2001) have noted that "10 ml to 100 ml water are incorporated during bathing (Johl *et al*. 1995)."



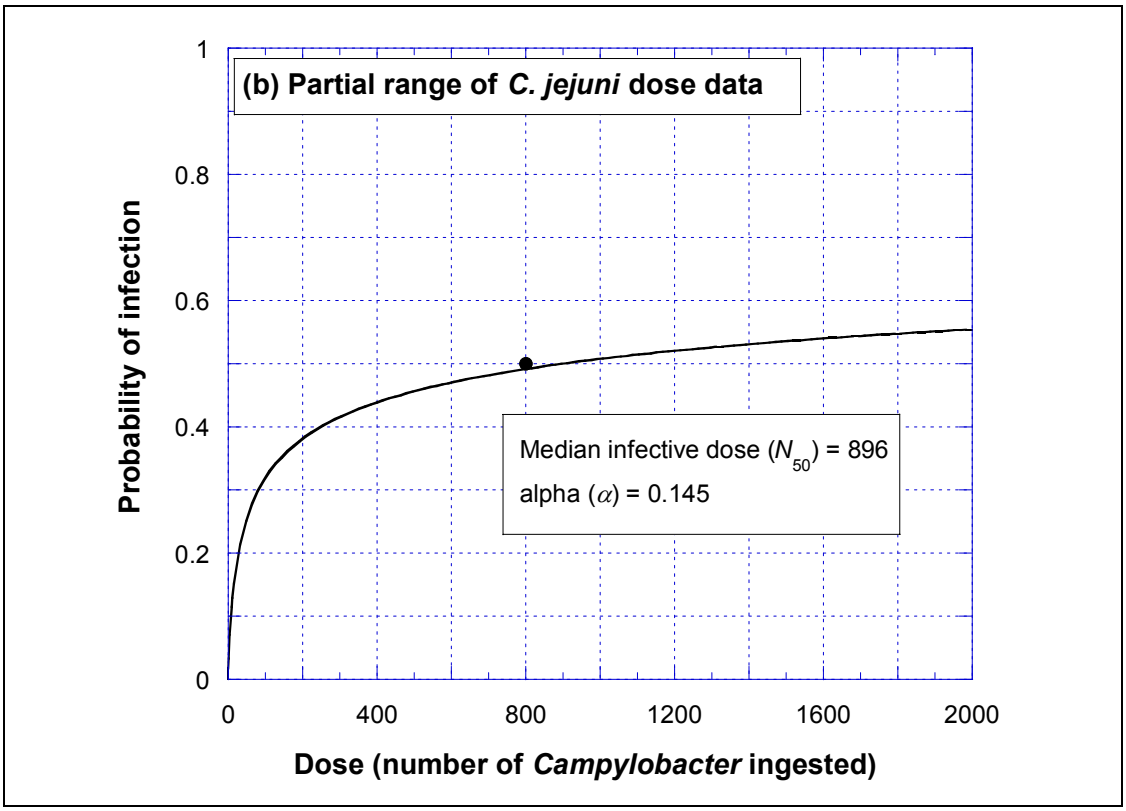


Figure A3.7.1 Dose-response curve for *Campylobacter jejuni* over full and partial range of dose data (parameter values from Table A3.7.1).[80]

[80] These α and N_{50} values have been obtained from a best fit to data in Black *et al.* (1988) who challenged 72 volunteers with *C. jejuni* A3249 (Penner serotype 27), obtained a few years earlier from a 16-year old boy with a sporadic infection after an outbreak at a camp in Connecticut (from which stock cultures had been maintained in glycerol at –70 °C). The doses administered (with infection proportions) were: 800 (5/10); 8,000 (6/10); 9×10^4 (11/13); 8×10^5 (8/11); 1×10^6 (15/19); 1×10^8 (9/9). Note that the lowest dose (800) isn't much less than the median infective dose (896 bacteria) calculated from the curve.

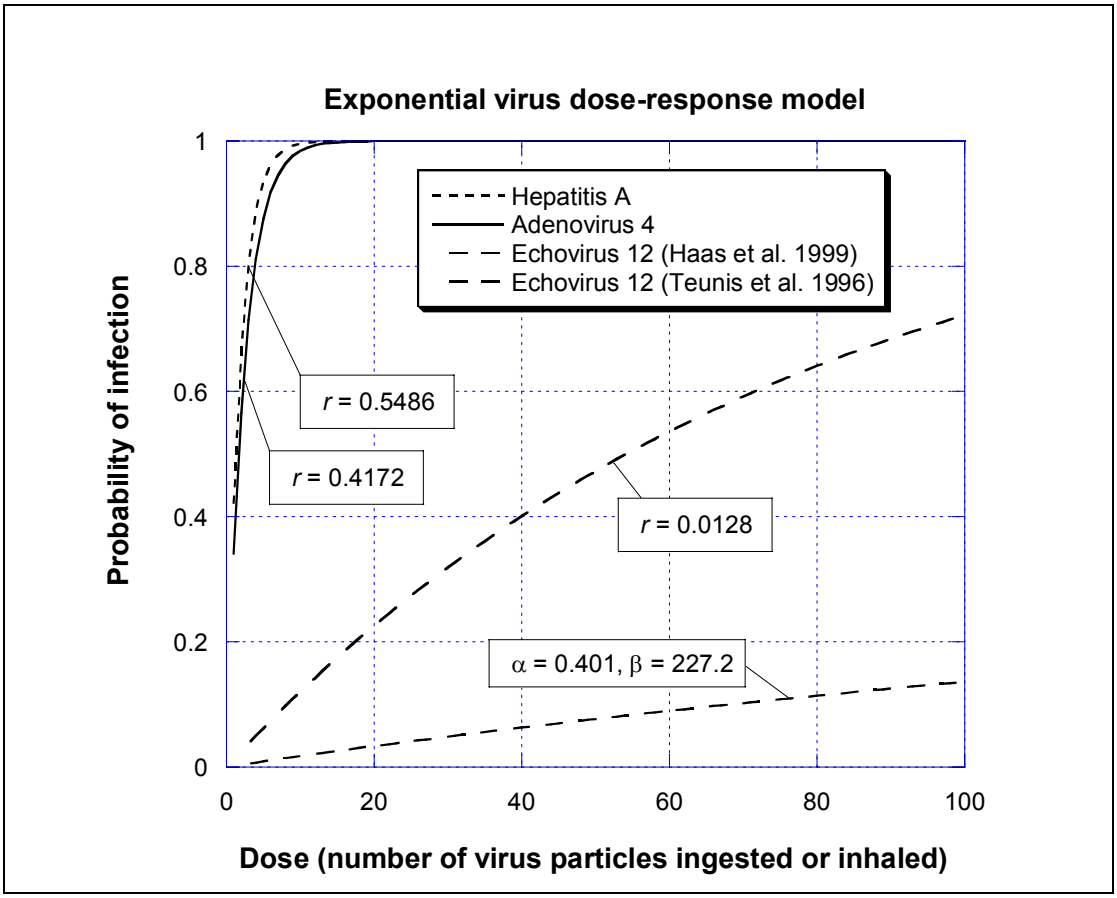


Figure A3.7.2 Dose–response curves for Hepatitis A, Adenovirus 4, and Echovirus 12 (taken as a typical enterovirus) and using parameter values from Table A3.7.1.[†]

[†]This graph shows that according to the fitted exponential model, the median infectious dose of adenovirus 4 is about 2 particles. This model is based on clinical trials on 15 subjects using aerosols containing adenovirus-4 particles (Couch *et al*. 1966). Nine subjects were exposed to small-particle aerosols (0.3–2.5 μ). Of these nine all 3 who received doses of 11 $TCID_{50}$ and all three who received doses of 5 $TCID_{50}$ became infected, but only one out of three became infected at a dose of 1 $TCID_{50}$. A further six subjects received a dose of 1,000 $TCID_{50}$ via a larger aerosol (15 μ), and all became infected. [Couch 1969 reports a further set of data that have not been used to date in identifying dose-response models: 21 subjects were exposed to small particle aerosols (1.5 μ). Of these all 4 who received a dose of 171 $TCID_{50}$ became infected, all 9 who received a dose of 5–11 $TCID_{50}$ became infected, 3 of the 5 who received a dose of 1–2 $TCID_{50}$ became infected and none of the three who received a dose of 0.1 $TCID_{50}$ became infected.]

The graph also shows the disparity between echovirus model parameters estimated by Teunis *et al*. (1996) from the data of Schiff *et al*. 1984a&b, and the fit reported by Haas (1999) for a different set of data (Akin 1981). The former set of data were obtained from a study of 149 subjects; the latter from a study of 60 subjects (as described by Haas 1983).

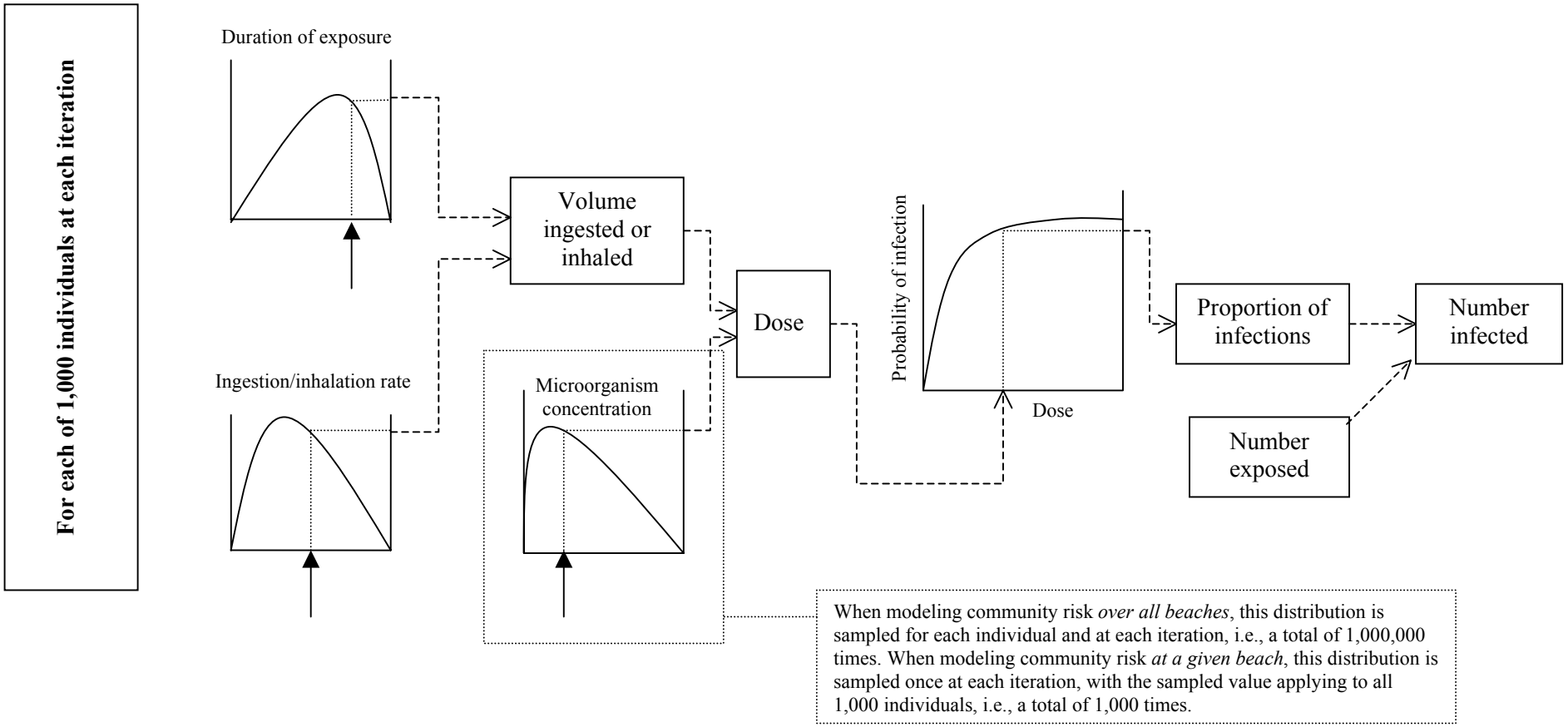


Figure A3.7.3 Monte Carlo simulation procedure for each of 1,000 individuals and 1,000 iterations. Information flows are shown as dashed lines. Random draws ("iterations") from three distributions are shown by solid vertical arrows; their position changes at each iteration. The shape of those three distributions governs the relative frequency of values obtained in the random draws, with values near the distribution's mode being rather more frequent than values in the tails.

The Monte Carlo approach proceeds by taking repeated random samplings from these distributions and in this way building up a distribution of infection probabilities. These random samplings are henceforth called "iterations"; we used 1,000.[81] This choice is simply conceptualised as 1,000 separate days, i.e., one iteration is one day's exposure (by 1,000 swimmers, water skiers etc.) to water that may contain illness-causing microorganisms. The final step in this analysis is to take those infection probabilities and calculate the distribution of infections per 1,000 population—health-related risks for recreational water exposure are traditionally reported per 1,000 people. We have used the "@RISK" software (Palisade Corporation 2000), being an "add-in" to Microsoft Excel spreadsheets.

Target populations

In performing these calculations one must be careful to consider what population is being modelled and how that is to be done.

The first point to be noted is that appropriate models require us to expose 1,000 people to random samples of water, with each person characterised by a random draw from the appropriate distributions for duration of exposure and for ingestion/inhalation rate of water. Random draws are also made from the distribution of microbiological contamination of the water. Therefore the model requires us to have 1,000 rows in the spreadsheet, with each row being "one person". Given that 1,000 iterations are to be made with those distributions, there will be one million sets of calculations (i.e., 1,000x1,000). Each simulation therefore takes some time.[82]

Second, we can rid the modelling of the difficulties posed by individuals' repeated exposures by assuming that there are a million people in total, each of whom has only one exposure.[83]

This is of course not a realistic scenario, but we have no data available to account for the possible effect of enhanced susceptibility to infection by reason of multiple exposures.

The last, and most important, question may be characterised as: "over what community is this risk to be spread?" Two cases may be considered here.

a) Community risk over all recreational sites. On each iteration the 1,000 people all go to 1,000 *different recreational sites*, the idea being that this represents the pattern of recreational water use by the public at large.[84] In this case each person is exposed to a different water body and so a random draw is made *for each person* from the distribution of microorganisms.
b) Community risk at a given recreational site. On any iteration the 1,000 people all go to the *same* site, the idea being that their spread of risks represents those faced by an individual *on a particular visit*. In this case we assume that they are all exposed to the *same* single random sample of microbiological concentration. Accordingly on each iteration *only one* random draw is made from the distribution of microorganisms to provide the microbiological data for all 1,000 people.[85] In terms of risks to be used in management this is arguably the risk of greater public interest.

Figure A3.7.3 contains a text box elaborating the procedure for making random draws of microorganism concentrations for these two cases.[86]

[81] Manly (1997) notes that 1,000 can be a desirable minimum number of iterations. We have performed some of these calculations using much higher iteration numbers and found results very similar to those reported here.
[82] About ½ hour on a Pentium 866 machine. In fact this calculation time can be considerably shortened in some *Campylobacter* cases, by sending only one person to the beach.
[83] Formally, this is the "ergodicity" assumption. That is, the result obtained by having a million people each with one exposure will be equivalent to having one thousand people with one thousand exposures.
[84] Ideally one would send 1,000 people to 1,000 sites 1,000 times, but a billion sets of calculations become prohibitive.
[85] That is, we are ignoring the temporal and spatial variation of microbiological quality at a recreational site, considering it to be much less than the variation one finds between sites. This seems a plausible assumption.
[86] Fortunately, issues to do with correlated random variables (Haas 1999) do not arise in this approach.

Sampling the *Campylobacter* distribution

Campylobacter have been enumerated by multiple tube fermentation techniques using a 3x3x3 dilution series, in which the following volumes were filtered: three of 100 mL, three of 10 mL, and three of 1 mL. These filtrates were then placed in nine tubes containing equal volumes of broth and incubated to see if fermentation occurred. At the end of the fermentation period the pattern of positive tubes was recorded. For example, "3-2-0" would mean that all three of the 100 mL filtrates tested positive for fermentation, two of the 10 mL filtrates tested positive, while none of the 1 mL filtrates proved positive. This pattern was then used to read the MPN ("Most Probable Number") of *Campylobacter* per 100 mL of sample from a "standard table". This table was constructed by considering all the probabilities of obtaining the given pattern of positive tubes for a range of candidate MPNs. The candidate having the highest "occurrence probability", i.e., the mode of the candidate MPN distribution, is then the MPN. This procedure means that MPNs are in a number system of a very different sort than the norm. In particular, they have two unusual properties:

- there are only a limited number of them (i.e., MPNs); most ordinary numbers are actually impossible,
- none have the same occurrence probability; some have very small occurrence probabilities.

In contrast, the numbers we usually deal with have these properties:

- all are possible,
- all have the same occurrence probability.[87]

On the top part of Figure A3.7.4 we show the frequency histogram of the 726 *Campylobacter* results obtained in the programme. The very jagged nature of the profile on that graph is the direct result of the properties listed above. The important consequence of these features is that one cannot sensibly fit a distribution to the data. If one tries, e.g., using the 'BestFit" feature in @RISK (Palisade Corp. 2000), one gets a single-parameter geometric distribution with parameter 0.0065713. This fit is heavily weighted by the values near zero and by all the impossible values (the curve fitting procedure "sees" these as zeroes). Such a fit is totally inadequate at large *Campylobacter* concentrations. That is, direct fitting of a parametric distribution to MPN data is totally inappropriate.

A simple alternative approach has been adopted to the sampling of this distribution, as depicted on the figure, consisting of four sequential steps:

- "Bin" the data according to some rational criterion, i.e., select the *Campylobacter* concentrations at the floor and roof of contiguous bins and count up all MPNs inside each bin, of whatever occurrence probability.
- Fit a distribution to the frequencies over the bins.[88]
- On any iteration, draw a random sample from that distribution to determine which bin is to be selected.
- Make a random draw from that bin using a uniform distribution between its floor and roof, so returning a *Campylobacter* concentration.

[87] That is, if the true value of some quantity is X, where X can be any (discrete or continuous) real number, then the most likely value from an unbiased estimator in sampling the distribution of X is in fact X. This property is not exhibited by MPNs. For example, if the true value of a *Campylobacter* were 28.8 per 100 mL, it is more likely that the pattern obtained in the fermentation tubes would correspond to a MPN of 23.7 (a 3-3-0 pattern, with occurrence probability 0.370) than 28.8 (a 3-2-3 pattern, with occurrence probability 0.002). (See Table A3.7.2 for occurrence probabilities.) More importantly, many numbers are actually impossible as an MPN value (e.g., 28.7, 28.69,…).

[88] We have used the "BestFit" feature in @RISK (Palisade Corp. 2000) for this distribution fitting. This package reports the fitting results for a number of discrete distributions and indicates which of them is the best. While there can be some advantage in using more advanced statistical packages for such fitting we did not do so in this case because the appropriate distributions (geometric and negative binomial) are catered for in the BestFit package.

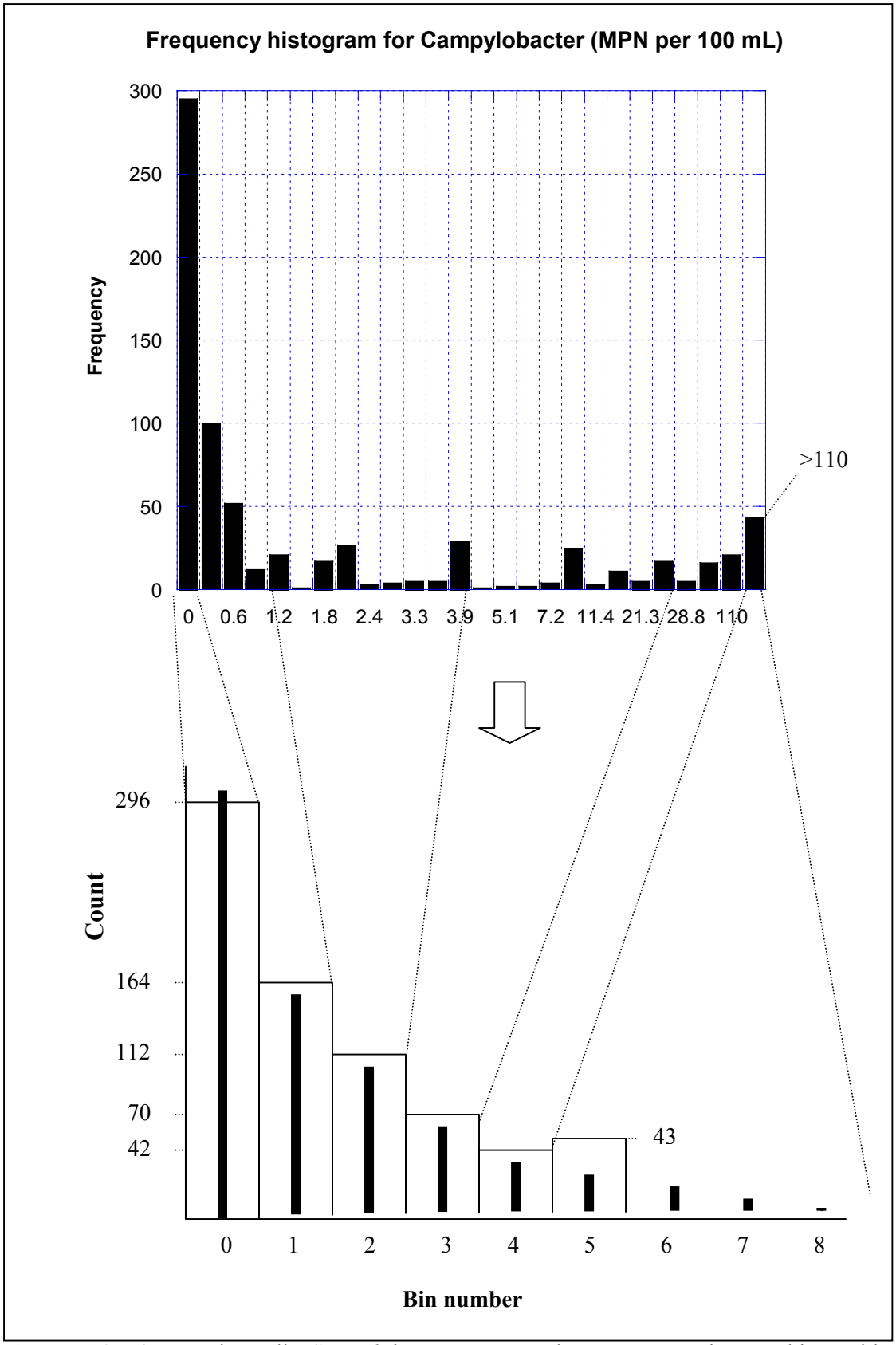


Figure A3.7.4 Mapping all *Campylobacter* MPN values onto contiguous bins with boundaries at 0, 1, 4, 25, 110 and >110 per 100 mL. The vertical bar in each bin is from the best fit geometric distribution (parameter = 0.42531). Not all bars in the bar graph (top of the figure) are labeled with their value (e.g., the bar between 1.8 and 2.4 is actually 2.1, as may be read from Table A3.7.2). Diagram for histogram of bins is not completely to scale.

Table A3.7.2 Most Probable Numbers of *Campylobacter* in the FMRP for all possible combinations of tubes showing a positive reaction in the series 3x100 mL, 3x10 mL, 3x1 mL.

Combination of positives	MPN /100 mL	Occur Prob[1]	FMRP Freq[2]	Combination of positives	MPN /100 mL	Occur Prob	FMRP Freq
0-0-0	<0.3	–	295	3-0-0	2.1	0.398	23
1-0-0	0.3	0.901	72	1-3-2		$7x10^{-7}$	2
0-1-0		0.090	14	1-2-3		$6x10^{-8}$	2
0-0-1		0.009	14				27
			100	2-2-1	2.4	0.002	2
2-0-0	0.6	0.541	27	2-1-2		$2x10^{-4}$	1
1-1-0		0.162	14	2-0-3		$2x10^{-6}$	0
1-0-1		0.016	3				3
0-2-0		0.005	2	2-3-0	2.7	0.003	2
0-1-1		0.002	4	1-3-3		$6x10^{-9}$	2
0-0-2		$5x10^{-4}$	2				4
			52	2-1-3	3	$2x10^{-6}$	0
1-2-0	0.9	0.015	5	2-3-1	3.3	$3x10^{-4}$	1
1-1-1		0.004	1	2-2-2		$7x10^{-5}$	4
0-2-1		0.001	2				5
0-3-0		$2x10^{-4}$	0	3-0-1	3.6	0.034	5
1-0-2		$1x10^{-4}$	2	3-1-0	3.9	0.400	26
0-1-2		$1x10^{-5}$	0	2-2-3		$9x10^{-7}$	3
0-0-3		$2x10^{-7}$	2				29
			12	2-3-2	4.2	$1x10^{-5}$	1
2-1-0	1.2	0.184	11	2-3-3	5.1	$2x10^{-7}$	2
2-0-1		0.018	2	3-0-2	6	0.002	2
1-3-0		$6x10^{-4}$	0	3-1-1	7.2	0.069	4
1-2-1		$5x10^{-4}$	2	3-2-0	9	0.339	25
1-1-2		$5x10^{-5}$	0	3-0-3	9.31	$4x10^{-5}$	0
0-3-1		$6x10^{-6}$	1	3-1-2	11.4	0.007	3
0-2-2		$2x10^{-6}$	1	3-2-1	14.7	0.129	11
1-0-3		$6x10^{-7}$	3	3-1-3	15.6	$3x10^{-4}$	0
0-1-3		$6x10^{-8}$	1	3-2-2	21.3	0.025	5
			21	3-3-0	23.7	0.370	17
0-3-2	1.5	$9x10^{-8}$	0	3-2-3	28.8	0.002	5
0-2-3		$9x10^{-9}$	1	3-3-1	45.9	0.430	16
			1	3-3-2	110	0.446	21
2-2-0	1.8	0.033	1	3-3-3	>110	–	43
2-1-1		0.009	9				
2-0-2		$3x10^{-4}$	3	TOTAL			726
1-3-1		$3x10^{-5}$	0				
1-2-2		$9x10^{-6}$	1				
1-1-3		$3x10^{-7}$	1				
0-3-3		$5x10^{-10}$	2				
			17				

[1] "Occur prob" is the occurrence probability (the probability of getting the stated pattern if the concentration of bacteria actually equals the MPN). "Freq" is the frequency of occurrence for the stated combination of positives in the FMRP. MPNs and occurrence probabilities were calculated using the computer program "XactMPN" (McBride in press).

[2] Some 2.8% of these data (i.e, 20/726) have an occurrence probability less than 1 in a million (41 have an occurrence probability < 10^{-3}). These samples have rather more positive results among the small volume tubes than among the larger volume tubes to an extent that is not plausibly just a chance event. One possible explanation is the possible presence of inhibitory material in the sample that gets diluted out in the small volume tubes. This could manifest itself in a clumping of such results for particular sites. Another possible explanation is laboratory errors that could be manifest in clumping of these results in time. There is no strong evidence of either effect (regarding the former, there is some clumping at site IA (Tukituki River at Black Bridge) in February and March 1999 and at site PA (Otaki River at Gauge in April 1999); regarding the latter there is a little clumping in January 1999). This matter would appear to be deserving of further investigation.

The criterion for bin selection was that each internal bin should contain two MPNs with high occurrence probabilities as being greater than 0.2 (these probabilities are given for all *Campylobacter* data on Table A3.7.2). This criterion was selected for operational convenience, as it gave very good fits in all cases. Bins were also defined at the extremes of the data (no tubes positive or all tubes positive, corresponding to zero and >110 per 100 mL). This procedure resulted in internal bin boundaries at 1, 4, 25 and 110 per 100 mL. These were used for analysis of the whole dataset, and also for analysis of data falling within each catchment type and with the bathing season only.

The fit of the bin frequencies to a geometric distribution is shown on the figure (this distribution's fit was the best of the various distributions examined in the "BestFit" package). Note that it includes bins 6, 7, 8… whereas our maximum bin number is 5. But this merely means that whenever a random draw is made from this distribution, any bin number greater than 4 signals that the *Campylobacter* concentration is in the ">110" category and the draw is to be made from Bin number 5. Because the sum of all the vertical bars in bins 5 and beyond matches the observed frequency in Bin number 5 closely, the frequency of drawing a value greater than 110 per 100 mL should be about right. Usually the best distribution fit was obtained using the simple geometric distribution. In one case (**B** catchments) the negative binomial distribution was best.[89] In order to make a draw from a uniform distribution in Bin Number 5 we need to define its roof. In consultation with microbiologists on the Risk Assessment Working Group this has been set at 2000 per 100 mL.

Note also that, from Table A3.7.1, the ID_{50} for *Campylobacter* is 896 particles, whereas the upper limit of detection in the FMRP is 110 particles per 100 mL and most water users would ingest or inhale less than 100 mL of water. However, this does not pose a large problem for the risk analysis, as the dose-response curve rises rather rapidly between 10 and 100 particles, and rather more slowly after that (as on Figure A3.7.1). It turns out that had the upper limit of the *Campylobacter* assay been in fact >1100 per 100 mL, not a great deal more information would be gained in the risk analysis (P_{inf} rises by only about 0.18 (from 0.33 to 0.51) in between 110 and 1100 bacteria).

The dose so-calculated can then be fed into the dose response equation (equation 2) to obtain a probability of infection (P_{inf}). Finally we take binomial samples from that series of infection probabilities for each individual (see equation 4 below), and count up the number of infections in the 1,000 cohort.[90]

So in summary the procedure is:

- For each person and at each iteration, draw the volume ingested or inhaled (v) and the bin number (do this only once per iteration in the case of the second target population—where all go to the same beach).
- Draw a sample of *Campylobacter* concentration from the bin, using the uniform distribution, and so compute each person's dose.
- Insert the doses into the dose-response equation (equation 3) to obtain a set of P_{inf} values.
- Take a binomial sample of all the infection probabilities to obtain the number infected:

$$\text{Infected (yes/no)} \sim \text{BINOMIAL}(1, P_{inf}) \qquad (4)$$

- Count up the number of infections in the 1,000 cohort.

Note that the *only* distributional assumption made in this procedure is invoked in the second step above, i.e., the *Campylobacter* distribution between the floor and ceiling of each bin is uniform.

[89] The geometric distribution is a special case of the negative binomial (i.e., where that distribution's integer-valued cluster parameter is unity).

[90] This step is necessary because we wish to compare *cases* of infection, not *probabilities* of infection. And even if those probabilities are very low we may, just by "bad luck" obtain the occasional case (statisticians call this "sampling error").

Sampling the virus distributions

Viral analyses are a special case, because their assay returns only positive or negative results (i.e., present/absent). For an ensemble of samples all we know from their assays is the proportion of them that are positive for the presence of virus material. We also note that such results, while obtained from 10 litre samples, *are in fact for one litre only* (i.e., the volume of each sample's pellet analysed in the laboratory test was adjusted so that it always represented one litre of the total sample from which the pellet was abstracted, G. Greening, ESR, Keneperu, pers. comm.). Furthermore, the volume of water ingested or inhaled is rather less then this 1 litre. Accordingly, in performing the Monte Carlo analyses we need some way of assigning a probability of infection to this small volume, given the overall proportion of positives in the one litre volume analysed. The only avenue open to us is to construct a probability distribution for the viral particles, as follows.

For target population (a)—1,000 people at different beaches

The viral results (per litre) typically show about 70% negatives. In this situation it is appropriate to use an "added-zeroes" distribution—because special account must be taken of this large clump of zero results. Assuming that any positive sample contains virus particles distributed at random, it is appropriate to use the Poisson with-added-zeros distribution (El-Shaarawi 1985, as also used by Teunis *et al*. 1997 in risk assessments for cysts in drinking water). This will require an unknown parameter to be stated—the estimated mean *of non-zero virus concentrations*, per litre. This will have to be based on expert judgement, and it seems best to do that using an upper and a lower bound. We again use a betaPERT distribution for λ, with minimum = 1, mode = 5 and maximum = 100 viruses per litre in positive samples (pers. comm. from virologists G. Lewis, Auckland University and G. Greening, ESR, Keneperu).

Let P be the proportion of positive results in V mL of sample [for example, P = 0.44 for adenoviruses in the bird (**B**) catchments, and V = 1,000, there being 1,000 mL in a one litre sample]. Then, from El-Sharaawi (1985)

$$P_{\text{none}} = \text{Prob(obtaining no viruses in volume } V) = (1-P) + Pe^{-\lambda} \qquad (5)$$

where λ is the expected mean of the non-zero counts in the volume V. This equation adds the probability of getting an "empty" sample ($1-P$) to the probability of getting a negative result from a contaminated sample ($Pe^{-\lambda}$). The complement of this probability is

$$P_{\text{some}} = \text{Prob (obtaining some viruses in volume } V) = P\left(1-e^{-\lambda}\right) \qquad (6)$$

(These two probabilities sum to unity, as required.) Also, we have

$$P_k = \text{Prob (obtaining } k \text{ viruses in volume } V) = P\frac{e^{-\lambda}\lambda^k}{k!}, \text{ where } k > 0. \qquad (7)^{91}$$

This last equation shows us that the probability mass function *for positive results* in volume V is just a weighted Poisson function, with weight P. Random sampling for values of k from such a distribution is simply coded in an @RISK model. We can then scale k down to the ingested/inhaled volume v, denoting this reduced count as c, using

$$c \sim \text{BINOMIAL}(k, v/V) \qquad (8)^{92}$$

(provided that $k > 0$). Here "~" means "is distributed as" and c is the virus count in the volume v when there are k viruses in the volume V. Random sampling from this binomial

[91] It may be simply shown that the sum of these probabilities over all the non-zero integer values of k is $P(1-e^{-\lambda})$, as required [because $\Sigma_{k=1,2,\ldots}(\lambda^k/k!) = e^{\lambda}-1$].

[92] More simply one could scale λ in (7) by v/V (P. Teunis RIVM, The Netherlands, pers. comm.). Similar results were obtained using this simplification.

distribution will return values of c from zero to k, with mode at $[v(k+1)/V]$, where "[]" denotes the "the integer part of".[93] This scheme guarantees that we will get a discrete number of viruses (including zero) in the ingested/inhaled volume v.

Finally, we can sample for either this result or for a zero result (we have to consider the possibility of obtaining a negative result also, i.e., no virus material found in the sample). This uses the discrete distribution:

$$\text{virus count in volume } v \text{ (i.e., the dose)} \sim \text{DISCRETE}\ (\{0, c\},\ \{P_{\text{none}}, P_{\text{some}}\}) \qquad (9)$$

This distribution simply assigns a probability P_{none} to a zero count and a probability P_{some} to the count c.

The dose so-calculated can then be fed into the dose response equation (equation 1) to obtain a probability of infection (P_{inf}). Finally we take binomial samples from that series of infection probabilities for each individual, using equation 4, and count up the number of infections in the 1,000 cohort.

So in summary the procedure is:

- For each person at each iteration draw the volume ingested or inhaled (v), and a value of λ—the expected mean of non-zero counts in samples of volume V.
- Draw a value of k (the number of particles in volume V) from the Poisson distribution with parameter λ.
- Draw a value of c (the virus count in volume v) using the binomial distribution (eq. 8).
- Draw the dose from equation 9.
- Insert the dose into the dose-response equation (equation 1) to obtain P_{inf}.
- Take a binomial sample of all the infection probabilities to obtain the number infected, using equation 4.

<u>For target population (b)—1,000 people at the same beach</u>
In this case a simplified procedure is followed, in which we first use equations 5 and 6 in a discrete distribution to make a random draw for whether the recreational site is contaminated or not. Values of λ and k are then calculated once per iteration and a non-zero dose is calculated from equation 9—but only if the recreational site is contaminated. In other respects the procedure is the same as above.

Dose-response curves
Figure A3.7.2 shows the marked difference in infection probabilities between adenovirus and enterovirus (i.e., Echovirus 12), using the parameters given by Haas *et al.* (1999) and shown on Table A3.7.1. In particular the ID_{50} has been reached once two adenoviruses are ingested, but, even under the Haas *et al*. (1999) dose response curve, this requires 55 enteroviruses.

Given the choice of two dose-response curves for enterovirus (see Figure A3.7.2), we have used the parameters reported by Haas *et al*. (1999), because it predicts higher risks—such a choice being consistent with precautionary public health practice.

One should also note that uncertainty about these curves has been quantified (Teunis & Havelaar 2000), in the form of likelihood-based confidence intervals. The uncertainty that these intervals describe should be included in future health risk modelling.

Results
The results for *Campylobacter*, adenovirus and enterovirus are shown on Tables A3.7.3–5.

[93] So for example if v = 85 mL, V = 1,000 mL, and k = 11, then the most common value of many draws made from this distribution will be c = 1. [Note that if $v(k+1)/V$ is in fact an integer, the distribution will have two modes, at $v(k+1)/V$ and at $v(k+1)/V - 1$.]

Results in all three cases show a median infection rate for a community at a given recreational site (left-hand panel of the Tables). However, the rates climb steeply at the higher percentiles. This is consistent with the idea that a majority of the time a given site is relatively uncontaminated, so nobody is infected; but occasionally the site can be rather contaminated, in which case many people *at that recreational site* can be infected.

In contrast, the right-hand panels of the Tables show appreciable median risks when the population-at-risk is spread over many recreational sites, but much less variation between the percentiles. This is consistent with the fact that at any one time a few sites will be relatively contaminated and so a few folk may become infected.

The results in these two panels demonstrate the need to clearly identify the population-at-risk.

The most useful statistics appear to be the median and 2.5%ile and 97.5%ile limits.[94]

We can summarise these results for the "all catchments" cases as follows

Infection	**Risk at a recreational site**	**Risk over many recreational sites**
Campylobacter infection	0 (0, 435)	41 (29, 54)
Adenovirus infection	0 (0, 402)	65 (50, 80)
Enterovirus infection	0 (0, 22)	3 (0, 7)

Numbers are mode (2.5%ile, 97.5%ile)

[94] Note that these limits are *not* confidence limits (which indicate the precision of estimation of a quantity such as a median). The limits given are percentiles and they indicate the spread of the results obtained—while not being influenced by extreme outliers. So 95% of all rates lie between these two limits (i.e., between the 2.5%ile and 97.5%ile). One should *not* use the range to indicate spread, because it is too influenced by chance events. For example, percentiles of the adenovirus infection rates calculated risks at a given recreational site for the **B** catchments (shown in the left-hand panel of Table A3.7.4) are generally higher than for all the other catchment types, but it has the lowest maximum. This is just unavoidable random noise arising from sampling infrequently in the tails of distributions.

Table A3.7.3 Calculated *Campylobacter* infection rates (per 1,000 people)

	Community risk at a given recreational site (1000 people at the same beach each day)							Community risk over all recreational sites (1000 people at 1000 different beaches each day)						
Percentile	**A**	**A(bs)**	**B**	**D**	**F**	**M**	**S**	**A**	**A(bs)**	**B**	**D**	**F**	**M**	**S**
Minimum	0	0	0	0	0	0	0	25	16	26	16	13	24	37
2.5%ile	0	0	0	0	0	0	0	29	21	33	22	18	30	41
5%ile	0	0	0	0	0	0	0	31	23	35	24	19	31	43
10%ile	0	0	0	0	0	0	0	33	25	37	26	21	34	46
15%ile	0	0	0	0	0	0	0	34	27	39	27	22	36	47
20%ile	0	0	0	0	0	0	0	36	28	40	28	23	37	49
25%ile	0	0	0	0	0	0	0	37	28	41	29	24	38	50
30%ile	0	0	0	0	0	0	0	38	29	42	30	25	39	51
35%ile	0	0	0	0	0	0	0	38	30	43	31	25	40	52
40%ile	0	0	0	0	0	0	0	39	30	44	31	26	41	53
45%ile	0	0	1	0	0	0	0	40	31	45	32	27	42	54
50%ile	0	0	2	0	0	0	0	41	32	45	33	27	43	55
55%ile	0	0	4	0	0	0	1	42	33	46	33	28	43	56
60%ile	1	0	9	0	0	1	3	43	33	47	34	28	44	57
65%ile	3	2	14	2	1	4	12	44	34	48	35	29	45	58
70%ile	9	4	19	4	3	9	18	44	35	48	36	30	46	59
75%ile	18	10	27	12	7	18	25	45	36	50	37	31	47	60
80%ile	26	19	57	19	17	28	63	46	37	51	38	32	48	61
85%ile	72	32	86	31	25	75	107	47	38	53	39	33	49	62
90%ile	131	91	139	93	73	146	211	49	39	54	40	34	50	64
95%ile	329	217	299	205	187	372	381	52	42	56	43	36	53	68
97.5%ile	435	413	430	398	391	441	438	54	44	59	45	38	55	70
Maximum	491	505	497	497	479	480	510	62	53	70	52	45	66	78
Statistic														
Mean	40.8	33.2	45.1	32.0	27.4	43.0	53.6	41.2	32.1	45.5	33.0	27.4	42.5	55.0
Mode	0	0	0	0	0	0	0	41	32	44	30	27	44	56
Std. Dev.	101.7	92.7	100.5	88.5	82.2	105.8	115.5	6.4	5.6	6.4	5.6	5.1	6.3	7.3
Skewness	3.04	3.51	2.94	3.52	3.94	2.93	2.42	0.16	0.18	0.15	0.20	0.24	0.08	0.17
Kurtosis	11.48	14.90	11.14	15.06	18.40	10.67	7.78	2.99	3.18	2.94	2.99	3.00	3.08	2.93

A = all catchments; **A(bs)** = all catchments, but in bathing season only; **B** = Bird catchments; **D** = dairy catchments; **F** = forestry/undeveloped catchments; **M** = municipal catchments; **S** = sheep catchments

Best fit bin distributions were: **A**, geometric (binomial parameter = 0.42531); **A(bs)**, geometric (binomial parameter = 0.46181); **B**, negative binomial (cluster parameter = 3, binomial parameter = 0.64804); **D**, geometric (binomial parameter = 0.45886); **F**, geometric (binomial parameter = 0.48441); **M**, geometric (binomial parameter = 0.42029); **S**, geometric (binomial parameter = 0.46181). Note that the geometric distribution is a special case of the negative binomial (with unit cluster parameter).

Table A3.7.4 Calculated adenovirus infection rates (per 1,000 people)

	Community risk at a given recreational site (1000 people at the same beach each day)							Community risk over all recreational sites (1000 people at 1000 different beaches each day)						
Percentile	**A**	**A(bs)**	**B**	**D**	**F**	**M**	**S**	**A**	**A(bs)**	**B**	**D**	**F**	**M**	**S**
Minimum	0	0	0	0	0	0	0	42	41	63	21	41	35	45
2.5%ile	0	0	0	0	0	0	0	50	48	70	30	48	43	59
5%ile	0	0	0	0	0	0	0	52	50	73	32	50	45	61
10%ile	0	0	0	0	0	0	0	55	53	77	34	53	47	64
15%ile	0	0	0	0	0	0	0	57	55	79	36	54	49	66
20%ile	0	0	0	0	0	0	0	58	56	81	37	56	51	68
25%ile	0	0	0	0	0	0	0	60	57	82	38	57	52	69
30%ile	0	0	0	0	0	0	0	61	58	84	39	58	53	70
35%ile	0	0	0	0	0	0	0	62	59	85	40	59	54	72
40%ile	0	0	0	0	0	0	0	63	61	86	41	60	55	72
45%ile	0	0	0	0	0	0	0	64	62	87	41	61	56	74
50%ile	0	0	0	0	0	0	0	65	62	88	42	62	57	75
55%ile	0	0	0	0	0	0	0	66	63	89	43	63	58	76
60%ile	0	0	34	0	0	0	0	67	64	90	44	64	59	77
65%ile	0	0	73	0	0	0	24	68	65	91	45	65	59	78
70%ile	32	16	112	0	21	0	70	69	67	93	46	66	61	79
75%ile	82	71	154	0	75	45	116	70	68	94	46	67	62	80
80%ile	144	120	193	22	133	107	158	71	69	96	47	69	62	81
85%ile	198	175	253	109	198	162	209	73	70	97	48	70	64	83
90%ile	264	250	304	201	243	246	274	74	73	100	50	72	66	85
95%ile	359	337	356	302	347	342	341	77	75	104	53	75	69	88
97.5%ile	402	398	404	384	415	423	400	80	78	106	55	77	71	91
Maximum	667	622	554	568	599	568	597	88	85	118	63	94	82	104
Statistic														
Mean	66.1	60.6	86.0	44.1	63.1	57.9	71.7	64.8	62.6	88.2	42.2	62.4	56.8	74.7
Mode	0	0	0	0	0	0	0	65	62	86	42	58	56	78
Std. Dev.	123.5	116.0	127.8	105.0	119.3	117.6	121.6	7.7	7.5	8.9	6.3	7.6	7.3	8.2
Skewness	2.00	2.04	1.38	2.57	1.99	2.16	1.75	0.00	0.04	0.06	0.03	0.10	0.12	0.10
Kurtosis	6.44	6.45	3.82	8.98	6.26	6.84	5.30	2.96	2.89	2.89	3.05	3.00	3.20	3.09

A = all catchments; **A(bs)** = all catchments, but in bathing season only; **B** = Bird catchments; **D** = dairy catchments; **F** = forestry/undeveloped catchments, **M** = municipal catchments; **S** = sheep catchments

Table A3.7.5 Calculated enterovirus infection rates (per 1,000 people)

	Community risk at a given recreational site (1000 people at the same beach each day)							Community risk over all recreational sites (1000 people at 1000 different beaches each day)						
Percentile	**A**	**A(bs)**	**B**	**D**	**F**	**M**	**S**	**A**	**A(bs)**	**B**	**D**	**F**	**M**	**S**
Minimum	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2.5%ile	0	0	0	0	0	0	0	0	0	0	1	0	0	0
5%ile	0	0	0	0	0	0	0	1	1	1	1	0	1	0
10%ile	0	0	0	0	0	0	0	1	1	1	1	1	1	1
15%ile	0	0	0	0	0	0	0	1	1	1	2	1	1	1
20%ile	0	0	0	0	0	0	0	2	2	2	2	1	2	1
25%ile	0	0	0	0	0	0	0	2	2	2	2	1	2	2
30%ile	0	0	0	0	0	0	0	2	2	2	3	2	2	2
35%ile	0	0	0	0	0	0	0	2	2	2	3	2	2	2
40%ile	0	0	0	0	0	0	0	3	3	3	3	2	3	2
45%ile	0	0	0	0	0	0	0	3	3	3	3	2	3	3
50%ile	0	0	0	0	0	0	0	3	3	3	3	2	3	3
55%ile	0	0	0	0	0	0	0	3	3	3	4	3	3	3
60%ile	0	0	0	0	0	0	0	3	3	4	4	3	3	3
65%ile	0	0	0	1	0	0	0	4	4	4	4	3	4	4
70%ile	1	1	2	3	0	1	0	4	4	4	4	3	4	4
75%ile	3	3	4	5	2	3	3	4	4	5	5	4	4	4
80%ile	6	6	6	7	4	5	5	5	5	5	5	4	5	4
85%ile	9	8	9	10	8	8	8	5	5	5	6	4	5	5
90%ile	11	12	13	13	10	12	11	6	6	6	6	5	5	5
95%ile	18	17	17	19	16	18	17	7	6	7	7	6	6	6
97.5%ile	22	20	21	24	21	22	21	7	7	7	8	6	7	7
Maximum	47	39	37	39	42	36	53	10	9	9	11	10	10	11
Statistic														
Mean	3.22	3.09	3.27	3.65	2.73	3.08	2.93	3.19	3.17	3.32	3.64	2.63	3.12	2.94
Mode	0	0	0	0	0	0	0	2	3	2	3	2	3	2
Std. Dev.	6.78	6.09	6.17	6.72	5.84	6.16	6.27	1.77	1.75	1.84	1.85	1.69	1.77	1.74
Skewness	2.89	2.39	2.18	2.19	2.66	2.37	2.91	0.62	0.50	0.40	0.59	0.72	0.50	0.50
Kurtosis	12.84	8.99	7.65	7.78	10.81	8.45	13.69	3.34	3.06	2.74	3.62	3.56	3.08	3.13

A = all catchments; **A(bs)** = all catchments, but in bathing season only; **B** = Bird catchments; **D** = dairy catchments; **F** = forestry/undeveloped catchments, **M** = municipal catchments; **S** = sheep catchments

GLOSSARY

Absolute risk	An incidence rate, usually expressed per 1,000 individuals.
Attack rate	The proportion of a disease-free population that becomes ill during a stated or implied period of risk.
Bacteriophage	A virus whose host cell required for replication is a bacterium.
Bather	Person who enters recreational water, not necessarily immersing their head.
Bias	A systematic error arising from faulty study design, data collection, analysis, interpretation, etc.
Campylobacter	Micro-aerophilic, thermophilic bacterial pathogen.
Campylobacteriosis	Enteritis caused by *Campylobacter* spp.
Categorical variable	A discrete variable (not continuous), denoting the category an event belongs to.
Clostridium perfringens	An anaerobic spore-forming sulphite reducing bacillus.
Cohort	A fixed population in which membership is permanent (in contrast to a *dynamic population*). Also defined as a group of persons who experience a certain event in a specified period of time (e.g., a birth cohort of babies born in 1990 in New Zealand).
Cohort study	A follow-up or longitudinal study that assesses exposure status before assessing outcome.
Coliphage	A bacteriophage whose host cell required for replication is a member of the coliform group.
Confidence intervals (95%)	Under the assumption of a given statistical model, the range of values for an estimated statistic constructed so that under repeated sampling its true value will lie within such *intervals* for 95% of the time. For a *particular* interval one cannot say with 95% confidence that the true value lies within the interval—unless one adopts a Bayesian approach, in which case a particular prior distribution will have been (probably unwittingly) adopted. This is then called a credibility interval. Confidence and credibility intervals usually have the same numerical limits if the prior distribution posits that all values are equally likely (and this may not be a tenable assumption).
Confounder	A variable that introduces bias into an association between a causal factor and an effect.
Confounding	A bias resulting from an unbalanced distribution of other causal factors, or markers for such factors, among people in different exposure categories.
Coprostanol	The principal human faecal sterol.
Cryptosporidium	A protozoa, *Cryptosporidium parvum* being a human pathogen.
Cryptosporidiosis	Diarrhoea caused by *Cryptosporidium parvum*
Cyst	Environmental resistant phase of a protozoon.
Effect modification	Modification by a second factor of the effect caused by a specified exposure.
Enterococci	A subgroup of faecal streptococci that are considered to be primarily of faecal origin. Enterococci are used widely as indicators of faecal pollution in marine waters.
Epidemiology	The study of the occurrence of disease, or other health-related variables, in human populations.
Exposure	Any characteristic or event that might cause or prevent disease.
Faecal coliforms	A subgroup of the coliform bacteria that can grow and ferment lactose with the production of acid and gas at 44.5°C and are generally regarded to be of faecal origin although some species can be of non-faecal origin. Also called thermotolerant coliforms.
Faecal sterols	Group of C27–C29 sterols (lipids) produced in the intestines of animals and humans by bacterial transformation of cholesterol and other dietary sterols.
Giardia	A protozoa, *Giardia lamblia* being a human pathogen.

Giardiasis	Diarrhoea caused by *Giardia lamblia*.
ID_{50}	Dose at which 50% of an exposed population become infected
Indicator, faecal	A microbiological organism (e.g., *E. coli*), or group of organisms (e.g., faecal coliforms), that is commonly and specifically associated with faecal material or illness risk among bathers.
Interaction	A joint effect, which is not constant among categories.
Linear correlation coefficient	A measure of the tendency for there to be a linear relationship between (X, Y) pairs of data. Measured by Pearson's coefficient (r), usually just called "correlation coefficient" (so the unwary may not realise that is measures only a linear relationship). See also “rank correlation coefficient”.
Mean, arithmetic	The sum of numbers divided by the number of numbers.
Mean, geometric	The n^{th} root of the product of n numbers; the same as the antilog of the mean of the logarithms. Tends to be a better measure of central tendency for skewed distributions, but is zero if any datum is zero. Estimates the geometric mean of a lognormal distribution, but with some bias (which can be corrected—see Gilbert, 1987).
Median	The middle value of n numbers (if n is even it is the arithmetic mean of the middle two numbers).
Oocyst	The resistant cyst form of protozoa such as *Cryptosporidium parvum* (the infective stage)
P-value (or p-value)	A probability calculated from a test statistic; it is the probability of obtaining data at least as extreme as has been obtained *if* the tested hypothesis were true.
Pathogen	A disease-causing or illness-causing microorganism (viral, bacterial, protozoan).
Percentiles	Values that divide the rank order of data into 100 equal parts.
Presumptive count	An estimate of the number of bacteria of a specific group in a sample, as revealed by an initial screening test. Requires further testing to be confirmed.
Prospective study	A study in which the disease events to be measured have not occurred when the study begins, and so study participants have no foreknowledge of their possible involvement. Both *follow-up studies* and *case-control studies* can be prospective with respect to their accumulation of cases.
Quartiles	Median of the bottom half of data (*lower* quartile), or of the top half of the data (*upper* quartile). That is, values that divides the rank order of data into fourths.
Rank correlation coefficient	A measure of the tendency for a Y value to increase, as it’s associated X value increases. It allows for monotonic non-linear relationships. Measured by Spearman's coefficient (r_s), sometimes called "Spearman's rho" (ρ).
Rate	In epidemiology "rate" has special usage; it is the frequency with which an event occurs in a defined population, at or over a specified period of time. A rate is therefore a ratio, and includes proportions.
Relative risk	The ratio of two incidence rates (strictly, *rate ratio* for person-time data); or the ratio of two cumulative incidences (strictly, *risk ratio* for count data). For "rare" diseases these ratios are approximately equal (hence, in common usage, both are referred to as relative risk). Loosely used to mean *odds ratio* in case-control studies since, for "rare" diseases, they may approximate one another.
Recreational Site	Area used for freshwater recreational activities i.e. lakeside beaches, riverbank picnic areas, stretches of river used for skiing
Recreational water	Water used for paddling (wading), swimming, and other water contact activities.

Risk	The probability that an individual will develop disease (or experience an event) in a specified period of time. Estimated by the cumulative incidence.
Risk factor	A cause, or indicator of a cause, of disease.
Salmonella	A pathogenic non-sporing bacteria.
Somatic	Of the body structure of an organism.
Spearman's correlation coefficient	See "Rank correlation coefficient".
Statistical significance	The condition in which the *p*-value is smaller than an *a priori* value (the significance level, α). If $p < \alpha$ one can then say that a result is unlikely to have arisen merely by chance, and that the result is "statistically significant".
Virus	The smallest microorganism comprising a central core of nucleic acid surrounded by an outer coat of protean structure. The nucleic acid of viruses is either deoxyribonucleic (DNA) or ribonucleic (RNA).
Zoonoses	Diseases of humans transmitted from animals.