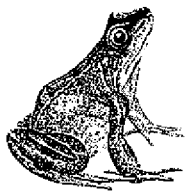
WRc

Development of a risk assessment model for infectivity of *Cryptosporidium* in drinking water

Final Report to the Department of the Environment

T84

DWI 4157/1
OCTOBER 1996



DEVELOPMENT OF A RISK ASSESSMENT MODEL FOR INFECTIVITY OF *CRYPTOSPORIDIUM* IN DRINKING WATER

Final Report to the Department of the Environment

Report No: DWI 4157/1

October 1996

Authors: P Gale, P van Dijk and E Carrington

Contract Manager: G Stanfield

Contract No: 10037

DoE Reference No: EPG 1/9/82

Contract Duration: January 1996 to March 1996

This report has the following distribution: 20 bound copies + 1 unbound external
4 internal

Any enquiries relating to this report should be referred to the Contract Manager at the following address:

WRc plc, Henley Road, Medmenham, Marlow, Buckinghamshire SL7 2HD.
Telephone: 01491 571 531

EXECUTIVE SUMMARY

Microbiological risk assessment models have been used to define targets for the degree of pathogen removal by drinking water treatment (Surface Water Treatment Rule) and to identify the maximum concentrations of pathogens in drinking water (Maximum Contaminant Levels) to ensure acceptable risk from waterborne pathogens to consumers.

In this contract, funded and managed by the Drinking Water Inspectorate, components have been assembled to develop a risk assessment model for *Cryptosporidium* in drinking water supplies. Existing risk assessment models developed for *Cryptosporidium*, *Giardia* and enteric viruses in drinking water rely on the assumption that pathogen organisms are randomly dispersed within 100 litre or 1,000 litre volumes. One manifestation of this assumption is that current models predict that consumers are either exposed to zero pathogens or to doses of just one pathogen. According to current models the probability of two or more pathogens being present in the same glass of water is negligible. Available evidence in the literature shows this assumption is not appropriate for drinking water.

Experiments performed in this contract have demonstrated that aerobic spores, which may serve as surrogates for *Cryptosporidium* oocysts, are not randomly dispersed in drinking water but are spatially associated to some degree, with some 100-ml subsamples containing considerably higher spore counts than others. It has been possible to accommodate some of the spatial variation by fitting the negative binomial statistical distribution. However, a small number of 100-ml samples contained very high counts which could not be accommodated. Although rare in occurrence, those high count samples could be of public health importance and need to be included in risk assessment models.

Overall, the data and most of the methodology to develop a model for assessing risks from waterborne cryptosporidiosis have been established in this contract. This is the first microbiological risk assessment model to account for some of the clustering of micro-organisms within drinking water samples. However, before a *Cryptosporidium* risk assessment model can be constructed further statistical consideration is needed to:-

1. define the effect of drinking water treatment on the dispersion of oocysts,

2. and to accommodate the few samples with very high counts.

From experiments with aerobic spores reported here, the data necessary to achieve the first of these two tasks are available. However, to characterise the rare occurrences of very high counts, much larger numbers of samples will be required.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

There is undisputed evidence that waterborne outbreaks of cryptosporidiosis can occur. In response to outbreaks in the United Kingdom and the United States of America a great deal of research has been carried out to gain a better understanding of this organism, particularly with regard to its transmission by the water route.

This research has highlighted the ability of oocysts of *Cryptosporidium* to survive in the aquatic environment and their tolerance of the disinfection processes used in conventional water treatment. A consequence of this resistance to disinfection is that oocysts of *Cryptosporidium* can still be present when coliform organisms, the traditional indicators of drinking water quality, have been inactivated.

To try and ensure freedom from risk of *Cryptosporidium*, other control strategies have had to be considered. Monitoring directly for the pathogen is of limited value because it will only be present in finished waters in extremely low concentrations and analytical procedures are both time-consuming expensive and imprecise. Risk assessment appears a promising alternative.

In this procedure data on the concentration of the micro-organism in the source water, the efficiency of treatment for its removal, dose-response data and exposure assessment based on quality of water consumed is used to model the likelihood of exposure to the pathogen.

Risk assessment models have been developed in the United States for organisms such as *Giardia* (Regli *et al.*, 1991) and enteric viruses (Haas *et al.* 1993). However, two shortcomings in these models have been identified. First they do not utilise data on the incidence of the pathogens in source waters even though monitoring at this point may be more certain and precise. Second, current models assume pathogens are randomly distributed within the drinking water supply and hence exposures to pathogens through drinking water may be modelled by the Poisson distribution (Regli *et al.* 1991; Haas *et al.* 1993; Rose and Gerba, 1991; Rose *et al.* 1991). There is considerable evidence, however, that bacteria, and therefore probably pathogens, are heterogeneously distributed across the drinking water supply (Maul *et al.* 1985; Christian and Pipes, Pipes *et al.* 1977).

Furthermore there is evidence for higher than expected variability in micro-organism counts even within 10-litre samples from the supply (Pipes *et al.* 1977). Thus micro-organisms appear to exhibit some degree of clustering within the drinking water supply. This is not considered by current risk assessment models. Those models may therefore miscalculate the proportions of the consumers exposed to different pathogen doses. Indeed, it was shown in "Review of Microbiological Risk Assessment and Drinking Water Supplies" (Report DWI 3940/1, Gale 1995) that the current virus model (Haas *et al.* 1993) predicts that even the most exposed individual never ingests a dose of more than one virus particle per day through drinking water.

It is conceivable, however, that through clustering, a small proportion of consumers could be exposed to high doses through drinking water. Furthermore, by ignoring clustering, current risk assessment models may over-predict the risks from more infectious agents such as rotavirus, but under-predict the risks from less infectious agents such as *Vibrio cholerae* and *Cryptosporidium*.

The Department of the Environment has contracted WRc to develop a risk assessment model for predicting the infectivity of *Cryptosporidium* in drinking water. This work has been managed through the Drinking Water Inspectorate. A risk assessment model would enable an estimation of whether conventionally treated drinking water makes an identifiable contribution to the background level of cryptosporidiosis in a community. In addition, risk assessment models would identify whether a perceived optimisation of a treatment regime represents a real reduction in risk of infection.

The approach is to develop a model for prediction of oocyst densities, based on water company source data and using data for efficiency of oocyst removal in drinking water treatment. The model is to use the United States Environmental Protection Agency's human infectious dose study (DuPont *et al.* 1995) and the Department's drinking water consumption study in developing estimates for exposure and risk of infection in different sub-groups of the exposed population.

1.1 Our approach

The central theme of our approach for developing a risk assessment model for *Cryptosporidium* in UK drinking water supplies is to consider the variability in micro-organism densities within drinking water.

Oocysts tend to be present in a high proportion of raw water samples, particularly river waters with upstream sewage works and agricultural inputs (Hutton *et al.* 1995). Conventional drinking water treatment removes a percentage (but not all) of those oocysts. For example, chemical coagulation followed by rapid gravity filtration removes 99 to 99.9% of oocysts. Our approach therefore is to attempt to model the distribution of oocyst densities in drinking water by simulating the effect of a 99 or 99.9% removal through treatment on the oocyst densities in the raw water. This part of our approach is not new and has been used in Dutch models for *Cryptosporidium* (Medema *et al.* 1995) and US models for *Giardia* (Rose and Gerba 1991). However, there are two problems which those approaches do not tackle and which we hope to overcome in this study. Those problems are:-

1. *Cryptosporidium* oocyst density data for surface waters are available only for much larger volumes (typically 100 to 1000 litres) than any one consumer would drink during a day (on average about 1 litre). Indeed, both the Dutch and US protozoan risk assessment models used source water density data from 100-litre sampling. The available density data, therefore, does not provide the information to enable risk assessment models to allocate the exposures of oocyst doses to the drinking water consumers. High resolution data at the 1-litre or even 100-ml volume sample is required to achieve this. For example, the highest density of *Cryptosporidium*

oocysts recorded in a drinking water sample in a US study (LeChevalier *et al.* 1991) was almost 50 per 100 litres. This density measurement alone, however, does not provide the information to predict exposures through drinking 2-litre volumes. According to US risk assessment models (Haas *et al.* 1993) the 50 oocysts would be randomly dispersed in that 100-litre volume and each person drinking 2 litres from that 100-litre volume would ingest, on average, one oocyst. However, oocysts might be more clustered in the supply, so that in an extreme case 49 consumers ingest zero oocysts and one consumer gets all 50 oocysts in his glass. The implications of this on the appropriateness of a risk assessment model are considerable as discussed in Gale 1995 (DWI 3940/1).

2. A second limitation to modelling the densities of oocysts in drinking water from source water density data and efficiency of removal by drinking water treatment is that treatment may actually promote clustering of the oocysts. This could occur when coagulant particles bearing concentrated oocysts break through into the supply. In addition back-wash events could promote the release of more concentrated oocysts into the supply. This is not considered by current approaches (Rose and Gerba, 1991; Medema *et al.* 1995) to protozoan risk assessment. Comparing statistical distributions of coliform densities from source water and its corresponding drinking water supply suggests that variability in coliform densities across the supply is greater than in the source water

From studies in the US (Pipes *et al.* 1977) it appears that even coliforms in well-mixed 10-litre volumes taken from the tap are not randomly dispersed but clustered to some degree. Therefore an important part of this contract will be to consider:-

a) how micro-organisms are clustered in drinking water supplies

b) how drinking water treatment changes the clustering in final water relative to that in the raw water.

Experiments with aerobic spores are described in Sections 7 and 8 and will enable the development of risk assessment models which take into account micro-organism clustering in drinking water at the resolution of volumes imbibed by drinking water consumers.

It is infeasible to perform pilot plant experiments with *Cryptosporidium* oocysts in this contract. Therefore, the variability between counts of a surrogate has been studied in water samples and that information used for modelling oocysts variability. Micro-organisms such as *Bacillus* spores are appropriate surrogates for oocysts because of their resistance to disinfection.

2. OBJECTIVES

The general objective of this contract is to develop a UK model for predicting the risk of infection from *Cryptosporidium* through drinking water. The model is to take into account micro-organism clustering in drinking water at the resolution of volumes imbibed by drinking water consumers.

The specific objectives in the tender document from the Drinking Water Inspectorate are:–

a) To predict concentrations of oocysts of *Cryptosporidium* in water supplies using source water concentration data and estimates for the efficiency of removal during drinking water treatment;

b) To estimate the distribution of oocysts in water supplies using methodology which takes account of the tendency of oocysts to associate in groups;

c) To obtain information on the relative proportions of different sections of the exposed community, including those sections considered to be at greater risk of infection;

d) To review information on the unboiled drinking water consumption habits of different sections of the exposed community, and the available evidence for infectivity of oocysts via drinking water consumption;

e) To develop and assess the feasibility of fully validating a model for the risk of acquiring cryptosporidiosis through the consumption of drinking water.

3. PROPERTIES OF *CRYPTOSPORIDIUM* OOCYSTS

Cryptosporidium oocysts are about 4 μm in diameter. Gregory *et al.* (1991) report a mean diameter of 4.15 μm with standard deviation of about 0.4. Results of Gregory *et al.* (1991) suggested that the oocysts were not aggregated in 2% NaCl but existed almost entirely as single oocysts. Oocysts have a low refractive index and it is possible that treated waters with low turbidities (e.g. 0.1 NTU) could contain high oocyst densities. Turbidity, however, is widely used for assessing the particulate impurities in water after filtration. Particle counting, although more expensive, would be a more reliable method of monitoring filtered water quality.

Electrical surface properties of oocysts may influence the efficiency of their removal by coagulation and filtration. Experimentally the zeta potential is calculated. The zeta potential is defined as the electric potential at the plane of shear between the moving oocyst and the suspending liquid. In tap water containing 2 mM Ca^{++} ions, zeta potentials for oocysts are in the range of -9 to -18 mV, i.e. oocysts have negatively charged surfaces. The surface charge of oocysts is pH-dependent, becoming less negative as the pH is lowered. For oocysts the isoelectric point (the pH at which net surface charge is neutral) is about pH 4. Typical water treatment additives (e.g. aluminium sulphate and the cationic polyelectrolyte, Percol 1697) also cause charge reversal on the oocyst surface.

3.1 Surrogates for *Cryptosporidium* oocysts

Gregory *et al.* (1991) have considered the particle properties of yeast cells and latex particles as potential surrogate particles for *Cryptosporidium* oocysts. The behaviour of yeast cells was found to be similar to that of oocysts. Latex particles have a higher electrophoretic mobility than oocysts in tap water. In the presence of alum the mobility of latex particles is similar to that of oocysts and yeast cells.

4. REVIEW THE AVAILABLE EVIDENCE FOR INFECTIVITY OF OOCYSTS VIA DRINKING WATER

4.1 Waterborne outbreaks

There is undisputed evidence that waterborne outbreaks of cryptosporidiosis can and do occur. Since the first identified waterborne outbreak arising from oocysts passing through water treatment plant in Carrollton, Georgia, in 1987 there have been a number of outbreaks in the USA, UK and elsewhere. The largest occurred in Milwaukee in 1993 when 400 000 persons were estimated to be affected, although less than 300 cases were laboratory-confirmed. The largest outbreak in the UK occurred in Swindon and Oxfordshire in 1989 and stimulated the establishment of the Group of Experts on Cryptosporidium in Water Supplies, led by the late Sir John Badenoch, and subsequently the National Cryptosporidium Research Programme. This Programme considerably extended the knowledge world-wide of this organism, particularly in relation to its spread by the water route.

An outbreak of disease is usually detected by a significant change in the number of cases above the background level and usually the source can be determined by epidemiological studies. However, the origin of the infection of the sporadic cases which form the background level is generally not known or investigated. There is a possibility, that the potable water supply may be a factor in the dissemination of low levels of infective material.

4.2 Oocysts in the environment

Cryptosporidiosis is widespread in domesticated animals and common in wild animals (Gregory 1990, Chambers *et al.* 1995). An infected calf may excrete up to 10^{10} oocysts per day over a short period and similar numbers will be excreted by infected humans. Most cattle infections occur in Spring, but there is recent evidence (Scott *et al.* 1994, Kemp *et al.* 1995) that the 'carrier state' can exist in adult cattle which may excrete the equivalent of 10^{7} oocysts per day through out the year.

Oocysts quickly lose their viability when retained in manure or pen leachate (Department of the Environment and Department of Health 1995). There is, at present, little evidence available on the survival of oocysts, or their viability, during passage through sewage treatment works.

Studies by Carrington and Ransome (1994) showed that oocysts exposed to the environment in river water during the warmer conditions of Spring and Summer deteriorated at a greater rate, both in numbers and viability, than those exposed at other times. Similar results have been reported by Robertson *et al.* (1992). In controlled conditions (Carrington and Ransome 1994) the number of oocysts declined by slightly

more than 1 $\log_{10}$ order over a 22-week period at consistent temperatures between 10°C and 30°C. The rate of decrease of viability was observed to be greater than the rate of decline of numbers, indicating that they do not disintegrate rapidly following loss of viability.

Observations in the laboratory suggest that oocysts tend to clump together when suspended in water.

In a survey of 3 rivers in the UK, 10 sites were sampled 3 times a week over a 12-15 month period (National Cryptosporidium Survey Group 1992). Positive samples were recovered from all sites and the oocyst concentrations ranged between 0.04 and 4 per litre (Table 4.1). Similar data has been reported from elsewhere in the world (Carrington and Smith 1995).

Table 4.1 Levels of oocysts observed in a survey of UK surface waters. (National Cryptosporidium Survey Group 1992).

Site	Number of samples	Number of positive samples	Range of positives (oocysts per litre)
A1	183	9	0.07-2.5
A2	192	8	0.10-4.0
B3	171	91	0.04-2.0
B4	160	90	0.06-1.59
B5	180	103	0.06-2.48
B6	180	79	0.06-3.0
C7	108	1	0.12
C8	106	7	0.07-0.73
C9	110	5	0.10-0.81
C10	106	5	0.10-2.75

4.3 Effects of water treatment

The Group of Experts on Cryptosporidium in Water concluded that effective treatment is the key factor in minimising the risk of waterborne cryptosporidiosis (Department for the Environment and the Department of Health 1995). Data from laboratory, pilot and full-scale plants demonstrate that effective removal of oocysts can be achieved by the two principal treatment processes; two-stage chemical treatment and slow sand filtration. A well operated plant will remove 99.9% or more of oocysts in the source water but this may be considerably reduced if the plant is operating at less than its design optimum.

Most disinfectants are ineffective, at concentrations that are acceptable for use in potable water, in destroying the viability of oocysts (Ransome *et al.* 1993).

4.4 Infective dose

Miller *et al.* (1990) suggested that less than 10 oocysts could cause infection in young macaque monkeys, but in another study, vervet monkeys were not infected by doses as high as 10^5 (Smith *et al* 1993 and 1994). Using lambs in conditions that simulated very young children, Blewett *et al.*(1992) used gnotobiotic lambs to try and simulate infectivity in young children. and concluded that the infective dose may be as low as one oocyst.

In the only study to use human volunteers, DuPont *et al.* (1995) found that 30 oocysts initiated infection in 20% of the volunteers exposed to this level. They calculated that the ID_{50} (dose required to infect 50% of the subjects) was 132 oocysts for adults with no serological evidence of previous infection with *Cryptosporidium.* The dose-response curve obtained from the study of DuPont et al. (1995) is used in the risk assessment model in Section 11.

Using iso-enzyme electrophoresis and polymerase chain reaction techniques McDonald and Awad-el-Kariem (1995) have demonstrated some strains of *Cryptosporidium* are specific to humans and others are specific to animals.

Robertson (1995) has argued that viability is not the same as infectivity and that the different tests for determining viability measure different aspects of the state. The convenient-to-use dye inclusion/exclusion test measures the integrity of the oocyst wall and the potential to move to the next development stage. The excystation test confirms the potential to release shizonts and thus, the possibility of initiating infection. Although the mouse infectivity test confirms the ability to initiate infection, it must be emphasised that the results may not have any relevance to infectivity in humans or other animals.

4.5 What is an outbreak

An epidemiologist would define an outbreak as two or more related cases. In practice an outbreak is identified when the number of cases is observed to rise above the background levels. For this observation to be made a number of events for each case must take place:

- the illness must be sufficiently persistent for the patient to approach his GP;
- the GP must be sufficiently concerned to submit a faeces sample to a laboratory;
- the laboratory practice and protocols must be such that they examine the faeces from that class of patient for protozoan parasites (Casemore and Roberts 1993); and
- positive results are reported to the appropriate body; this is not compulsory in England and Wales.

An outbreak situation may be easy to identify when the patients have activities in common, as in a party of school children visiting a farm, but it is considerably more difficult with sporadic cases with no obvious link, such as may occur from a low level of contamination of a water supply.

The occurrence of cryptosporidiosis in recent years in England and Wales for humans, cattle and sheep, from all sources of infection, is indicated in Table 4.2. For humans the numbers indicated the number of individual cases reported, but the animal figures refer to

the number of herds involved. For the reasons outlined above these figures can only be taken as a guide to the size of the problem.

Table 4.2 Occurrence of cryptosporidiosis in recent years in humans, cattle and sheep in England and Wales.

Year	Human	Cattle	Sheep
1989	7904	925	159
1990	4682	865	102
1991	5165	918	108
1992	5211	1101	95
1993	4819		
1994	4424		

Sources: PHLS Communicable Disease Surveillance Centre and Central Veterinary Laboratory.

It has been suggested that the first indications of the outbreak in Milwaukee in 1993 (*vide supra*) was when the local pharmacies ran out mixtures for the self-treatment of diarrhoea.

4.6 Susceptible populations

The portion of the population most susceptible to infection will be those with an incomplete immune system, and will include the young, the elderly, those persons using immuno-supressant drugs (eg transplant patients) and those with disease conditions that affect the immune system, such as AIDS.

Since reporting of HIV infection began in 1982, a total of 25 276 cases had been reported in the UK up to late 1995. Of these 11 814 had developed AIDS of whom 8112 had died (Communicable Disease Report 1995).

The Chief Medical Officers' Expert Group on AIDS has advised those persons with CD4 counts below 200 μl^{-1} to boil any water before drinking it, and also advised others with HIV infection to seek professional advice on the subject (Department of the Environment and the Department of Health 1995).

After comparing the prevalence of cryptosporidiosis in AIDS patients in an area of Los Angeles County for the four years before the installation of filtration equipment at the local water works with the four years after the installation and with the prevalence in a neighbouring community whose water was filtered for the whole period, Sorvillo *et al.*(1994) concluded that municipal drinking water was not an important risk factor for cryptosporidiosis in those communities.

Between April 1990 and June 1991 the Communicable Disease Control Centre (CDSC) interviewed 191 patients who had experienced cryptosporidiosis.. Their results (Marshall, Holmes and Stanwell-Smith 1991) are summarised in Table 4.3.

Table 4.3 The interview response of 191 cryptosporidiosis patients. (Marshall, Holmes and Stanwell-Smith 1991)

Criteria	Percent of Respondents	Criteria	Percent of Respondents
Case status		*Tapwater consumption*	
Primary case	71	None	14
Secondary case	27	1-2 glasses d^{-1}	35
		3-4 glasses d^{-1}	20
Gender		5-6 glasses d^{-1}	13
Male	54	> 6 glasses d^{-1}	3
Female	46		
		Foods consumed	
Age		Unpasteurised milk	7
< 1year	9	Raw meat	10
1-4 years	55	Sausages	73
5-14 years	9	Cold meat	76
15-44 years	19	Hamburgers	40
45-65 years	3		
> 65 years	5	*Animal contact*	
		Pets, sick or young	17
Symptoms		Farm	20
Diarrhoea	98		
Abdominal pain	58	*Travel*	
Adults - 80%		Nights away	35
Children - 50%		Local	12
Vomiting	51	UK	16
Adults - 30%		Abroad	6
Children - 59%			
Loss of appetite	74		

The study confirmed that young children are more susceptible than other age grouping. The CDSC suggest that not only do they represent the primary cases but also pass the infection to their susceptible peers at school or playgroup. A higher proportion of cases lived in households where there was a child under 5 years than that shown in the 1981 census for all households in England and Wales. The authors also point out that parental concern ensures that children with diarrhoea are more likely to see a doctor than are adults.

It is likely that after infection the antibodies persist in the individual for considerable periods. In discussing 'herd' immunity, (i.e. the level of immunity within a community), Casemore (1995) points out that if the levels were moderate to high this would be indicated by relatively low levels of apparent infection in adults. This in turn means that the young within a community are more susceptible than adults. However, he argues that it is possible for symptomless infection to occur in the partially immune. The hypothesis of moderate to high levels of herd immunity would also explain why travellers appear to be susceptible to infection, as they are in contact with an antigenic type which differs from their previous experience.

4.7 Comment

The available evidence does not dispute the possibility that sporadic cases of cryptosporidiosis, both temporal and geographical, which may or may not be diagnosed, may arise from small numbers of infective oocysts entering the water supply.

The Group of Experts on Cryptosporidium in Water Supplies in their second report (Department of the Environment and the Department of Health 1995) comment "The absence of *Cryptosporidium* oocysts in drinking water can never be guaranteed ..."

5. CRYPTOSPORIDIUM DENSITY DATA FOR RAW WATERS IN THE UK

Data are available for large volume samples in the region of 10 - 1,000 litres. Pathogens densities are higher in raw waters than treated waters and fewer zeros are recorded. *Giardia* and *Cryptosporidium* have been found in 53.9% and 60.2%, respectively of surface water samples of collected in the US between 1988 and 1993 (LeChevallier and Norton, 1995). Densities of *Cryptosporidium* oocysts from rivers with agricultural drainage both in the US and the UK can be regarded as log-normally distributed (Hutton et al 1995) as are *Giardia* cyst densities from three pristine rivers in the Pacific Northwest (Ongerth 1989).

5.1 Published density data for raw waters.

Densities of both *Cryptosporidium* oocysts and *Giardia* cysts are reported to be lognormally distributed in raw river waters (Ongerth 1989; Hutton *et al.* 1995). Parameters defining log-normal distributions were estimated from the normal plots in Hutton *et al.* (1995) and are presented in Table 5.1, together with the 2.5% and 97.5% percentiles.

Table 5.1 Parameters defining log-normal distributions for *Cryptosporidium* oocyst densities in five surface waters (Hutton *et al.* 1995).

Water Supply	Log_{10} density (counts per 100 litres)		Percentiles (counts per 100 litres)		
	Mean (μ)	Std. Deviation (σ)	2.5%	50%	97.5%
River A (agricultural drainage)	2.929	0.575	63	850	11380
River B (rural domestic drainage)	2.212	0.424	24	163	1104
Farmoor Reservoir[a].	-0.5	1.575	0.0003	0.31	386
River Tillingbourne[a]	0.0	1.88	0.0002	1	4839
Protected Watershed (*Giardia*)[a]	0.425	1.00	0.029	2.7	243

[a]To estimate $\mu$and σ for pathogen densities in raw source water, best fit lines defining log-normal distributions (Hutton *et al* 1995) were extrapolated because 55 - 90% of samples were below limits of detection and contained zero counts.

5.2 UK data available

Levels of *Cryptosporidium* oocysts in four UK surface water sites B3, B4, B5 and C8 (Table 4.1) in the survey of the National *Cryptosporidium* Survey Group 1992 are plotted as log-normal distributions (Figs). Data appear to be approximately log-normal and parameters are presented in Table 5.2.

Table 5.2 **Parameters defining log-normal distributions for *Cryptosporidium* oocyst densities in four surface water sites in the UK. Data from National Cryptosporidium Survey Group.**

Surface Water Site	Log_{10} density (counts per 100 litre)		Percentiles (counts per 100 litre)		
	Mean (μ)	Std. Deviation (σ)	2.5%	50%	97.5%
B3	0.624	0.833	0.406	6.80	113.9
B4	1.247	0.402	2.87	17.6	108.4
B5	1.106	0.528	0.023	3.4	495.9
C8	-0.335	0.942	0.0066	0.46	32.5

To estimate $\mu$and σ for pathogen densities at each site, best fit lines defining log-normal distributions in Figure 5.1 to 5.4 were extrapolated because 55 - 90% of samples were below limits of detection and contained zero counts.

Figure 5.1 **Log-normal distribution for Cryptosporidium oocyst densities in a UK surface water (Site B3) studied by National Cryptosporidium Survey Group 1992.**

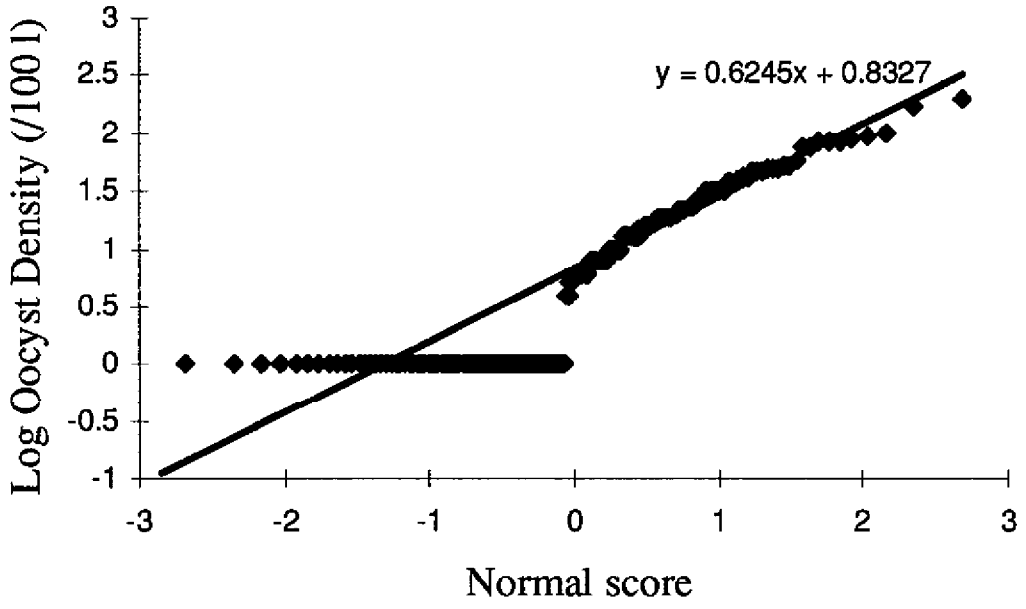


Figure 5.2 **Log-normal distribution for Cryptosporidium oocyst densities in a UK surface water (Site B4) studied by National Cryptosporidium Survey Group 1992.**

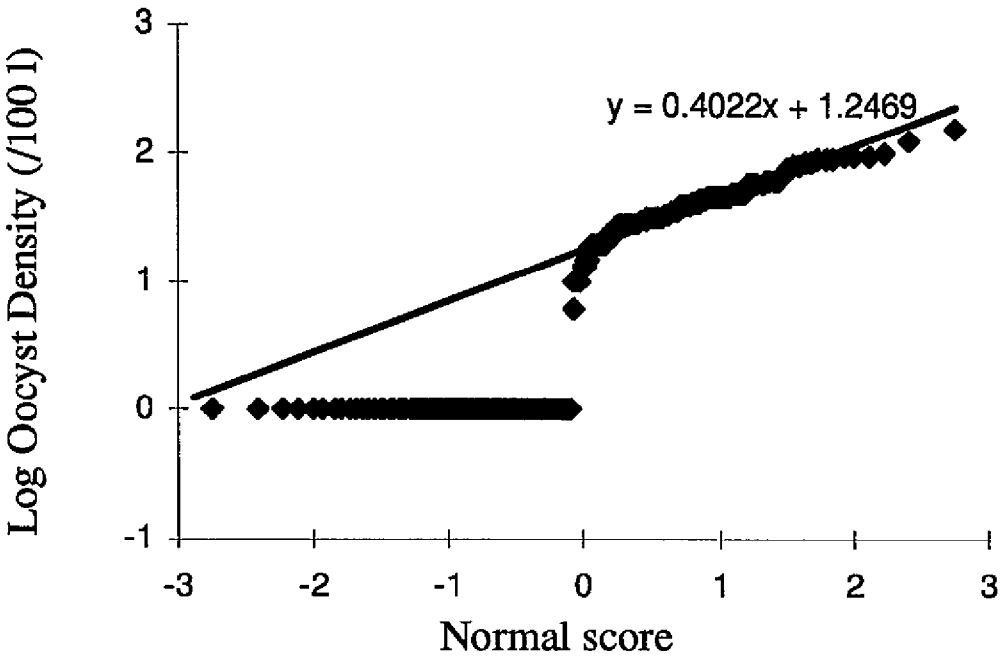


Figure 5.3 **Log-normal distribution for Cryptosporidium oocyst densities in a UK surface water (Site B5) studied by National Cryptosporidium Survey Group 1992.**

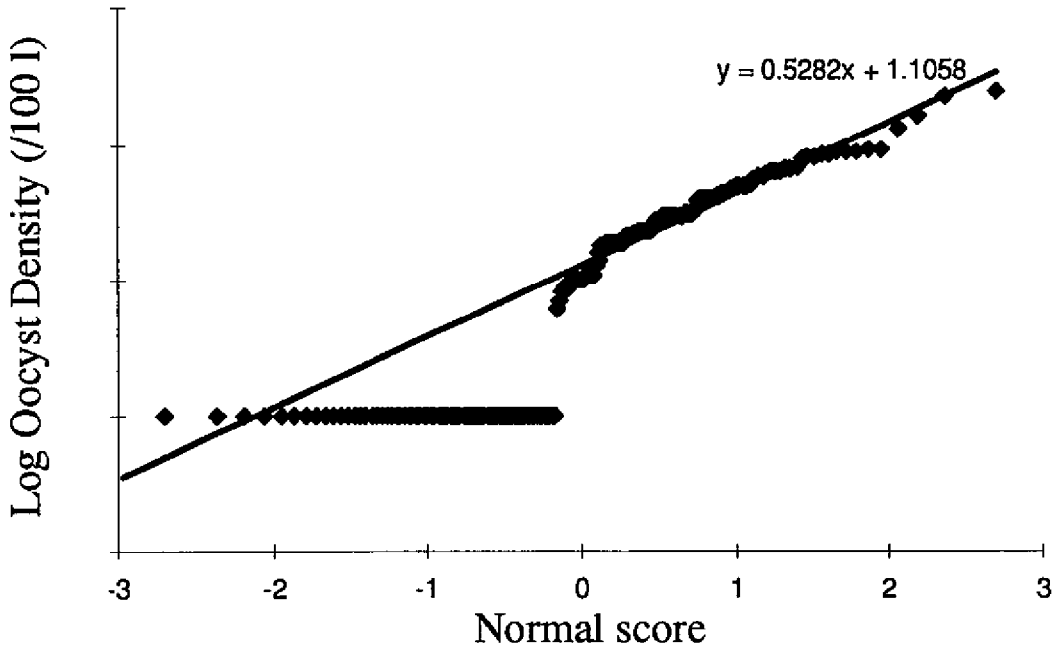
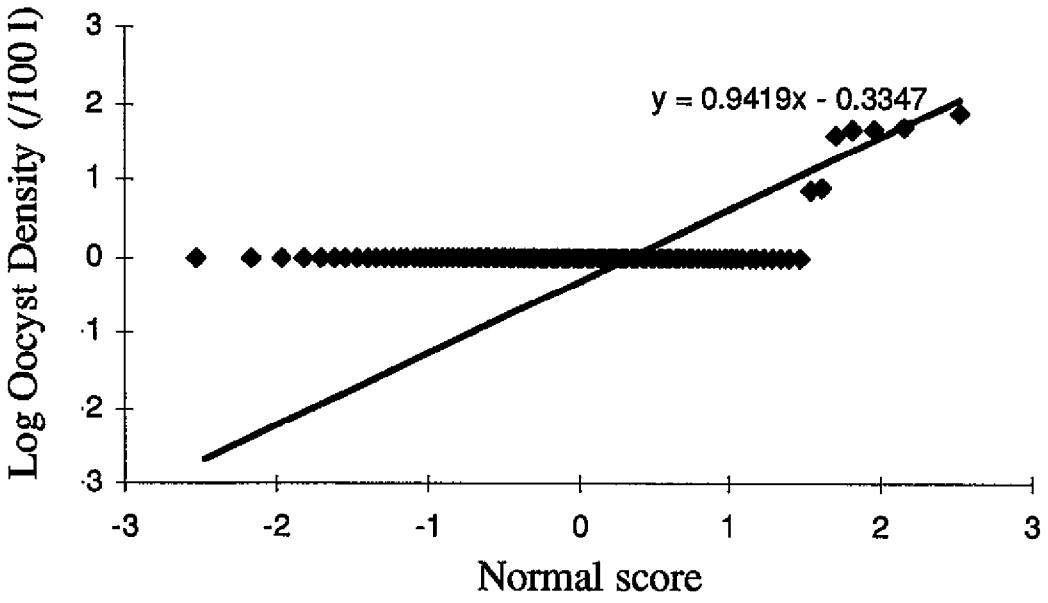


Figure 5.4 **Log-normal distribution for Cryptosporidium oocyst densities in a UK surface water (Site C8) studied by National Cryptosporidium Survey Group 1992.**



5.3 Variation in densities between large volume samples in raw water

From Table 5.1 and Table 5.2, it would appear that there is considerable variation in oocyst densities between large volume samples (100 litres) taken from surface waters. Furthermore some 100 litre surface water samples contain very high counts of *Cryptosporidium* oocysts. In Site B5 in the UK some 2.5% of 100 litre volumes contained over 500 oocysts. Much higher densities are apparent from the data reported in Hutton et al 1995 (Table 5.1) with some samples containing over 10,000 oocysts.

5.4 Modelling raw water data as negative binomial distributions

In the *Cryptosporidium* risk assessment model (Medema *et al.* 1995) the variability between oocyst counts in large volume raw water samples was modelled with the negative binomial distribution. This approach is not appropriate for the *Cryptosporidium* oocyst data in the four UK surface water sites B3, B4, B5 and C8 (Table 4.1 and Table 5.2) because different volumes of raw water were analysed. This may also be a limitation of the model of Medema *et al.* 1995.

6. REVIEW UNBOILED TAP WATER CONSUMPTION IN THE UK FOR DIFFERENT SECTIONS OF THE COMMUNITY

6.1 Total tap water consumption

Means for total tap water consumption in the 1978 (Hopkin and Ellis 1980) and 1995 (M.E.L. Research 1995) UK surveys are compared in Table 6.1. Since statistical distributions for tap water consumption are highly skewed and suggested to be log-normal (Roseberry and Burmaster 1992), the arithmetic mean is not a particularly good estimate of central tendency, being influenced by a few values with large magnitudes. The median would have been a more appropriate estimate.

Table 6.1 Mean tap water intakes (litre person^{-1} d^{-1}) in UK.

Year	Male	Female
1978	1.07 in adults	
1995	1.127	1.149

6.2 Boiled vs. unboiled water

Most waterborne pathogens are inactivated by boiling. Therefore, unlike risk assessment models for chemical parameters, microbiological risk assessment models should only consider the consumption of unboiled tap water. Current microbiological risk assessment models developed by the US EPA (Regli *et al.* 1991; Haas *et al.* 1993) use a value of 2 l person^{-1} d^{-1}. The *Cryptosporidium* model of Medema *et al.* (1995) uses data on log-normal frequency distributions from Roseberry and Burmaster (1992). However, both approaches appear not to distinguish unboiled and boiled tap water consumption. Indeed, tap water intake in the analysis by Roseberry and Burmaster (1992) is defined as the sum of water drunk directly as a beverage, and water added to foods and beverages during preparation. Thus, the microbiological risks from current microbiological risk assessment models are likely to be overestimated to some degree by not considering the proportion of pathogens inactivated through boiling.

The proportion of tap water consumed after boiling will vary between countries and within a country depending on culture. In the UK, for example, almost 80% of tap water is consumed as coffee or tea (Table 6.2). For this reason, risk assessment models will need to be calibrated for use within a particular country.

Table 6.2 How tap water is consumed in the UK in 1995. Taken from National Drinking Water Survey 1995 by M.E.L. Research.

Type of Drink	% of all tap consumption
Water	9.2
Tea	49.2
Coffee	29.1
Hot drinks other	7.5
Other sources	1.8

According to the 1995 survey it would appear that, in total, only 12.4% of tap water in the UK is consumed unboiled (Table 6.2). In the 1978 survey (Hopkin and Ellis, 1980) the mean for hot tap water-based drink consumption (0.95 l person^{-1} d^{-1}) was considerably greater than for cold tap water-based drink consumption (0.12 l person^{-1} d^{-1}). This would suggest that in 1978, on average, only 11% of tap water was consumed unboiled in the UK. Between 1978 and 1995, therefore, it would appear that the proportion of tap water boiled has not changed much (Table 6.3).

Table 6.3 Unboiled tap water consumption in the UK national surveys.

Survey	% tap water unboiled
WRc 1978	11.2%
M.E.L. 1995	12.4%

6.3 The statistical distribution for consumption of unboiled tap water in the UK.

A frequency distribution for consumption of cold water drinks in the UK 1978 study is shown on page 27 of Hopkin and Ellis (1980). The distribution is skewed and could be approximately log-normal as reported in the US study (Roseberry and Burmaster, 1992). The original data are no longer available (Ellis, personal communication). However, using the frequencies for each volume class consumed estimates for the geometric mean and variance may be obtained (Table 6.4).

Table 6.4 Log-normal distribution for consumption of unboiled tap water.

Study	μ	σ	Percentiles (litres head^{-1} day^{-1})		
	$\log_{10}$	$\log_{10}$	2.5%	50%	97.5%
UK 1978	-0.76	0.345	0.036	0.173	0.823
US 20 - 65 age group[a]	-.091	0.212	0.047	0.123	0.322

[a]summing 11.2% of tap water is consumed unboiled (Table 6.3).

Table 6.4 also shows the different percentiles for the log-normal distributions for unboiled tap water consumption calculated from the formulae:-

2.5 pecentile = $10^{(\mu - 1.96\sigma)}$.
50 pecentile (geometric mean) = $10^{(\mu)}$.
97.5 pecentile = $10^{(\mu + 1.96\sigma)}$.

In the UK in 1978, 95% of the population drink between 36 and 823 ml of unboiled tap water per person per day. Parameters describing a log-normal distribution for unboiled drinking water consumption in the UK are also presented in Table 6.4 based on the consumption for the 20 - 65 year age group in the US study (Roseberry and Burmaster 1992) and assuming that UK consumers drink 11.2% of the tap water as unboiled (Table 6.3). Using this approach the 97.5% for volume of unboiled tap water consumed is much lower than reported in Hopkin and Ellis (1980).

6.4 Age variation in unboiled tap water consumption

In the 1978 UK survey (Hopkin and Ellis, 1980) beverage consumption was categorised according to different age groups. In general, total tap water intake increased until about the age of 18 (Table 6.5). The mean volume of total tap water boiled for tea or coffee also increased with age. Mean intakes of tap water not consumed as tea and coffee are calculated in Table 6.5. In general in males over the age of 18, less than 20% (by volume) of tap water is consumed unboiled. In male children under the age of 11 years, the intake of unboiled tap water is much greater, accounting for over 50% of total tap water consumed. This corresponds with higher consumptions of cordial by children.

Table 6.5 **Summary of mean tap-water based intakes (l person^{-1} day^{-1}) for males in different age groups. Data from 1978 UK study (Hopkin and Ellis 1980).**

Age Group	Total Tap Water	Tea & Coffee	Unboiled (%)*	Cordial
1-4	0.477	0.227	0.250 (52%)	0.188
5-11	0.550	0.241	0.309 (56%)	0.133
12-17	0.805	0.512	0.293 (36%)	0.090
18-30	1.006	0.816	0.190 (19%)	0.054
31-54	1.201	1.064	0.137 (11%)	0.020
55+	1.133	1.012	0.121 (11%)	0.020
***calculated as tea and coffee intake subtracted from total tap water intake**				

7. DISPERSION OF MICRO-ORGANISMS IN DRINKING WATER SUPPLIES

7.1 Pathogen density data needed for risk assessment

The dispersion of micro-organisms in drinking water is a key factor in determining the pathogen doses to which drinking water consumers are exposed. Pathogen density data in drinking water supplies are typically reported for sample volumes in the order of 100 - 1,000 litres (LeChevallier *et al.* 1991; Payment *et al.* 1995). The problem is that volumes of unboiled tap water consumed per person per day are at least two orders of magnitude smaller (Table 6.4).

Therefore to provide a complete picture of the variation in pathogen densities in the drinking water supply, information is needed on the variability in pathogen densities at two levels:-

1. Between 100 or 1,000 litre volumes

2. Within 100 or 1,000 litre at the volumes imbibed daily(30-800 ml of unboiled water)

7.2 Variation in pathogen density between 100 litre or 1,000 litre volumes

For the purpose of predicting pathogen densities in drinking water from raw water data it is assumed that the variation in counts between 100 litre or 1000 litre volumes is the same as that in the raw water (5.1 and Table 5.2).

7.3 Spatial variation in pathogen counts within 100 litre volumes

7.3.1 Published evidence for coliform clustering in 10 l samples

Pipes *et al.* (1977) studied the variability in coliform counts between 100-ml subsamples taken from a 10-litre tap water sample. Even in such a small volume the coliforms were not randomly dispersed but spatially associated to some degree. Indeed superimposing a Poisson distribution on the frequency histogram for observed counts showed that a higher number of 0/100-ml counts were observed than expected and, at the other extreme, a few samples contained much higher counts than expected. Those experiments suggest that the β-Poisson or exponential dose-response curves used in current risk assessment models (Regli *et al.* 1991; Haas *et al.* 1993; Medema *et al.* 1995) are not appropriate.

7.3.2 Evidence for pathogen clustering in drinking water

Pathogen densities measured in the drinking water supply in 100 - 1,000 l volumes also show variation, with most samples registering zero, but some with high counts (Table 1). Densities of *Cryptosporidium* oocysts and *Giardia* cysts in raw waters show large variation and appear to be log-normally distributed (Hutton *et al.* 1995). Drinking water treatment while diminishing pathogen densities by several log orders may promote clustering further. Thus, chemical coagulation concentrates *Cryptosporidium* oocysts into floc particles some of which, perhaps laden with oocysts or other pathogens, may break-through into supply. Filtration also concentrates pathogens. The practice of recycling filter backwash water with highly concentrated *Cryptosporidium* oocysts may overwhelm filtration resulting in break-through of oocysts into finished water (Lisle and Rose 1995). Such break-through events are more likely to release pathogens as concentrated clusters into the supply than as a dilute, homogeneous stream.

Rogers and Keevil (1995) have demonstrated that *C. parvum* oocysts attach in clusters to biofilms in aquatic environments and are able to persist and survive for several weeks within that biofilm community at 20°C. Rogers and Keevil (1995) suggest that subsequent sloughing of biofilm may release a small but effective dose of oocysts into the water system.

Cryptosporidium data

Evidence that treatment does increase the variation in densities of *Cryptosporidium* oocysts between samples is provided by large scale pilot plant trials (Hall *et al.* 1995). Raw waters were spiked with relatively homogeneous oocyst densities of 300 - 800 l^{-1}. After coagulation and filtration oocysts were not detected in half of the 1,000 l samples analysed but numbered between 10 and 600 in the remaining 1,000 l samples, suggesting that oocyst densities varied over 600-fold in the treated waters.

The authors note that there was no indication that the higher oocyst levels were associated with greater spiked concentrations or occurred at times of high filtered water turbidity. However, most of the higher values for oocyst concentration occurred when there were indications of operational abnormalities on the plant, such as increased clarified water turbidity or high metal ion coagulant concentration in the filtered water, usually when the raw water turbidity was higher than usual.

7.4 Experiments to determine statistical distribution of micro-organisms within 100 litre volume drinking water samples

7.4.1 Objectives

The aims of the experiments described in this section are two fold:-

- To determine how micro-organisms are dispersed in large volume samples at the resolution of volumes imbibed by drinking water consumers.

- To determine how drinking water treatment changes micro-organisms dispersion relative to that in the raw water.

This information will facilitate the development of risk assessment models by enabling:-

- the allocation of pathogen doses to consumers through drinking water, and

- the prediction of pathogen densities in drinking water from raw water data and known removal efficiencies by drinking water treatment

7.4.2 Methods

Pilot plant experiments involving the spiking of raw waters with *Cryptosporidium* oocysts were not feasible because the intensity of sampling and analysis required would have been prohibitively expensive. In addition the precision of the analysis technique for *Cryptosporidium* is low placing further uncertainty in the results. Therefore, aerobic spores were used as surrogates. The drinking water treatment pilot plant, at WRc Medmenham, could not be used because of cold weather damage. A drinking water treatment works was therefore chosen where raw water samples and samples after PAC precipitation and rapid gravity filtration could be collected. The sampling points used were clean metal taps. Preliminary microbiological analyses showed no coliform bacteria in the filtered waters, reflecting low numbers in the raw waters. Suitable numbers of aerobic were present in the filtered waters. Experiments therefore focused on monitoring spores in a 100-l volume of raw and filtered samples.

Sampling protocol.

Volumes of 100-l were sub-sampled from stainless steel taps at the treatment works. The flow rate was adjusted to just over 4 l min^{-1}. Exact flow rates were measured before and after taking the 100-l volume because the flow rate dropped by 10 - 20%. An average flow rate was used to calculate the total volume of water. The tap was allowed to run for 5 min before commencing sampling. Sub-samples of 150-ml were taken every 30 s for filtered water and every minute for raw water. Sub-samples were collected in sterilin bottles and stored at 4°C in the dark during transit. Analyses were performed within 2 h.

Analyses of aerobic spores.

To break up clumps of micro-organisms, Tween detergent (0.1% v/v) was added to the samples with shaking 30 m prior to analysis. Samples were heated to 80°C in a water bath and maintained at the temperature for 12 minutes to destroy all vegetative organisms. For the 20 raw water samples, serial dilutions of 0.9 ml and 0.1 ml were analysed by

membrane filtration (18 h; 30°C). For the 40 filtered water samples, 100 ml and 10 ml volumes were analysed.

Counting colonies

For these experiments it was crucial to ensure that any variability between plates was real and not an artefact of the enumeration method. Therefore, consistency between plates in the counting of colonies was of critical importance. In initial experiments it was difficult to be sure that the same counting procedure was being used for all plates. A major problem was the small size of many colonies after 18 h incubation. While some colonies were visible to the naked eye, many more smaller colonies became apparent using a magnifying glass. In the first two sampling exercises, colonies counts were only made at 18 hr. In subsequent experiments it was found that colonies were more easy to detect if 2,3,5-triphenyl-tetrazolium chloride (TTC) was added under the filter 1 hr before counting. TTC stains metabolically active colonies red. A further 18 hr incubation at 30°C with TTC produced larger colonies which greatly facilitated counting although on occasions plates would be unreadable because some colonies became too large. There was a need to define a standardised method for enumeration to ensure consistency.

The standardised method is as follows. Colonies were counted after 18 h (with TTC) and again at 36 h at 30°C. All red-stained colonies apparent with a magnifying glass were counted. Colonies were counted independently by two staff. Counts at 36 h were used.

7.4.3 Results

Sampling exercises were performed on seven days during February. Means and variances for sub-samples in 100-l filtered and raw water volumes are shown in Table 7.1 and Table 7.2. The method of counting is also indicated.

Table 7.1 **Statistics for aerobic spores in 100-ml volume samples from 100-l filtered water samples after PAC coagulation and rapid gravity filtration.**

	100 ml			10 ml			Counting
	mean	variance	n	mean	variance	n	Efficiency
1	40.07	189.1	41	4.27	6.60	41	not standardised
2	43.4	66.9	38	2.36	2.29	38	not standardised
3	65.1	522.2	36	9.54	16.7	33	standardised but only counted by one member of staff
4	65.1	169.4	40	3.15	4.18	40	not standardised
5	56.3	5705	39	9.35	98.6	40	standardised
6	75.7	2293	39	7.55	24.25	39	standardised
7	most counts over 200			30.8	44.1	32	standardised

Table 7.2 **Statistics for aerobic spores in 100-ml volume samples from 100-l raw water samples.**

	0.9 ml			0.1 ml			
	mean	variance	n	mean	variance	n	Counting
1							not standardised
2	15.8	19.2	16	2.1	1.58	16	not standardised
3	24.6	3.4	8	3.57	5.70	19	standardised but only counted by one member of staff
4	27.1	52.6	18	3.85	4.76	20	standardised
5	22.9	12.7	17	2.3	1.8	20	standardised
6	43.6	60.5	17	6.44	6.37	18	standardised

The standardised method, which worked well with good agreement between the two staff, was only implemented in the last three sampling exercises (Table 7.1 and Table 7.2). Counts recorded in Experiment 3 were determined by the standardised method, but only enumerated by one member of staff. For the purposes of the development of risk assessment models it is only acceptable to rely on data obtained during exercises 3, 5, 6, and 7. To give some indication of whether the counts are Poisson distributed, the dispersion index is calculated as:-

sample variance x number of samples (n) / sample mean

The dispersion index is then compared with the tabulated value of χ^2 with (n - 1) degrees of freedom. The hypothesis that spore counts are Poisson distributed would be rejected if the dispersion index exceeded the upper 5% point of χ^2 in the statistical tables. Dispersion indices together with the level of statistical significance at which the Poisson model is rejected are presented for sampling exercises 3, 5, 6, and 7 in Table 7.3.

Table 7.3 Dispersion indices (degrees of freedom) for aerobic spores in raw and filtered water.

	Filtered		Raw	
	100 ml	10 ml	0.9 ml	0.1 ml
3	288.5 (35)**	57.9 (32)***	1 (7)	30.2 (18) NS
5	3951 (38)***	421.8 (38)***	34.9 (17)**	24.7 (19) NS
6	1181 (38) ***	125 (38)***	9.4 (16) NS	15.7 (19) NS
7	nd	45.7 (31)*	23.6 (16) NS	17.8 (17) NS

nd, Not determined
Statistical significance at which hypothesis of Poisson distribution is rejected is indicated; * 5%; ** 1%; *** 0.1%; NS, Not significant

The main finding is that in all the filtered waters examined the distribution of aerobic spores in 100-ml and 10-ml sub-samples from a 100 litre volume were not randomly distributed (Table 7.3). In contrast, counts were not incompatible with Poisson distributions in six of the eight analyses performed on the raw water samples. This would suggest that treatment could increase the variability of counts within 100-l volumes relative to the Poisson distribution.

The most striking example was found in sampling exercise 5. Variation in spore counts between 100-ml sub-samples are compared in 100 l volumes of raw and filtered water in Figure 7.1 and Figure 7.2. In both figures densities (/100-ml) are plotted as logarithms on Normal probability plots. The raw water densities were calculated as the sum of the 0.9 and 0.1 ml counts multiplied by 100. In Figure 7.1, filtered water densities are calculated

as the 10-ml volume multiplied by 10. In Figure 7.2, counts form 100-ml volumes are used.

It is apparent from Figure 7.1 that spore counts vary more within the filtered water than with in the raw water. Thus, while the average count from the 40 x 10-ml filtered samples was 9.35, three samples contained no spores and four samples contained more than 30 spores. This is considerable variation.

Inspection of the 100-ml counts (Figure 7.2) shows some evidence of spore clustering in the filtered water. Thus, although variation between counts in 34 of the 39 filtered samples appeared similar to that in the raw waters, the other five samples showed higher counts. Indeed the highest count, (488) was almost nine-fold the mean (Table 7.1).

Plotting the spores counts (per 100-ml) as a function of the cumulative volume of the filtered water 'run' from the 100 litre sample for the 10 ml samples in Exercise 5 showed that the high counts did not all occur in the same region of the 100 litre sample (Figure 7.3). Spikes are apparent.

Evidence of clustering was also apparent in Exercise 6. In particular, one sample contained much higher spore counts than the others (Table 7.4).

Table 7.4 Aerobic spore counts were much higher in one sample than the other 39 filtered water samples. Data for Experiment 6.

Volume	Mean*	Range*	Sample 7
100 ml	68.2	49 - 96	359
10 ml	6.9	1 - 14	33

*excluding sample 7

7.4.4 Discussion

Seven exercises have now been conducted to establish the variation in counts of micro-organism in large volume samples at the resolution of volumes imbibed daily by drinking water consumers. Aerobic spores were used as surrogates for *Cryptosporidium* oocysts. Consistency in the counting method used for colonies after membrane filtration is of critical importance. A 'standardised' counting method has therefore been developed and was applied for four of the seven exercises.

The effect of drinking water treatment (poly aluminium chloride coagulation followed by rapid gravity filtration) on the dispersion of aerobic spores was measured in samples taken from a local treatment works. Spore counts were enumerated by membrane filtration in 100 ml sub-samples from 100 litre raw and treated water volumes. Results obtained so far suggest that aerobic spores are not randomly (Poisson) distributed within large volume samples of the treated water (as is assumed in current microbiological risk assessment

models). Indeed, counts showed greater variation than expected in the Poisson model. There was evidence to suggest that in some cases spore counts within large volume raw water samples were randomly (Poisson) distributed. Thus drinking water treatment, while decreasing spore counts, appeared to increase the variability in counts between samples.

There was evidence to suggest that treatment promoted the clustering of spores. Indeed, spore counts were four to eight-fold higher than the mean count in a small proportion of sub-samples taken from the 100 l treated water volume. In one 100 l volume analysed, the clusters appeared not to be spatially related. No evidence for clustering was found in the raw waters. The actual degree of clustering may be greater because only 4% (40 x 100-ml) of each 100 l volume sample was analysed. Furthermore there is evidence to suggest that biofilms in the drinking water distribution itself may further promote the clustering of *Cryptosporidium* oocysts.

These findings are of considerable importance for the development of microbiological risk assessment models for drinking water.

Figure 7.1 **Variation in spore counts between 150 ml sub-samples in 100 l volumes of raw and treated water (10 ml counts) from Experiment 5**

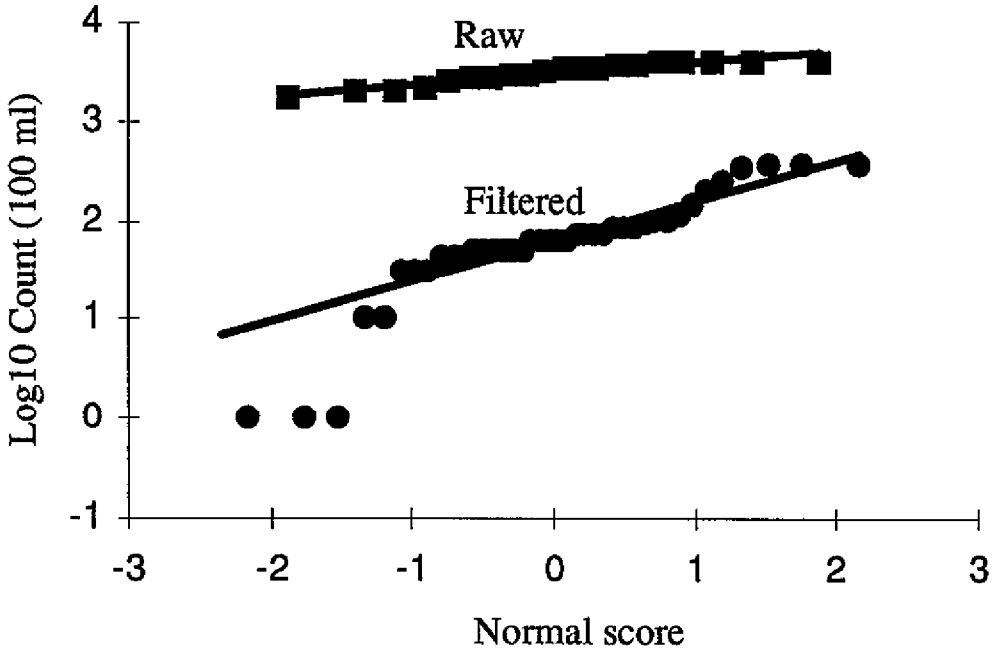


Figure 7.2 **Variation in spore counts between 150 ml subsamples in 100 l volumes of raw and treated water (100 ml counts) from Experiment 5**

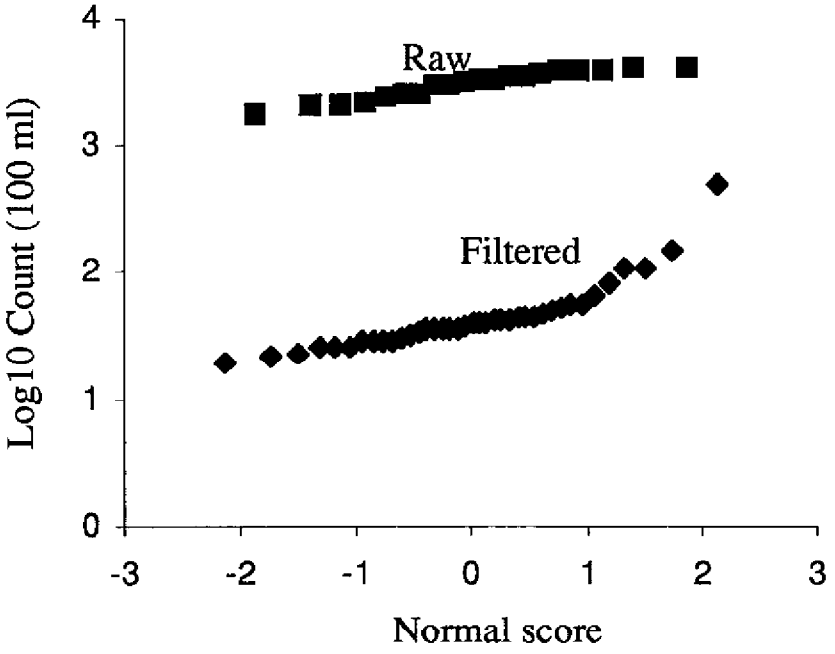
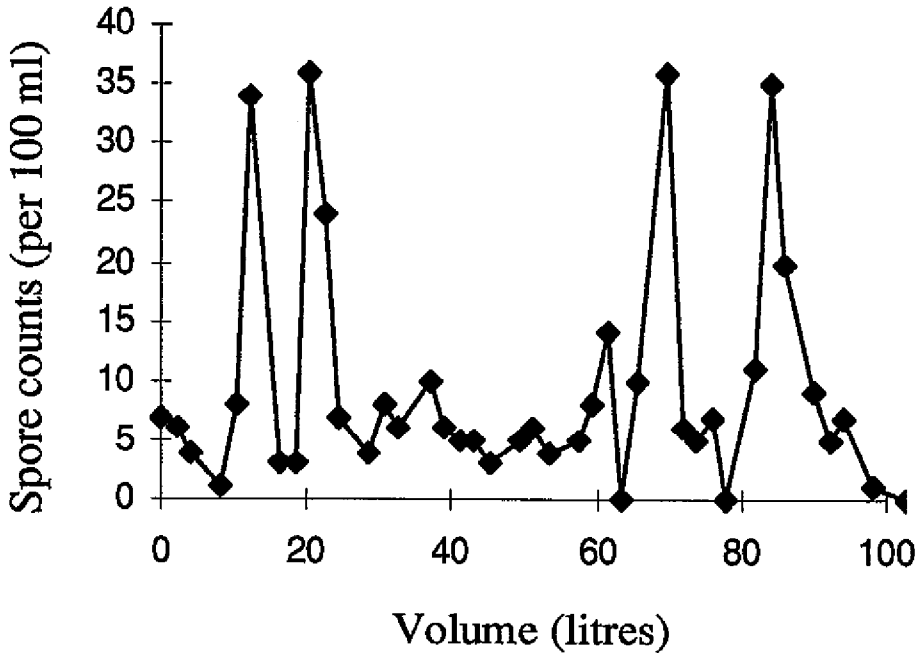


Figure 7.3 **Variation in counts of aerobic spores between 150 ml subsamples from a 100 l sample of treated water. Data for 10 ml sub-volumes in Experiment 5.**



8. MODELLING BACTERIAL COUNTS IN DRINKING WATER BEFORE AND AFTER TREATMENT

8.1 Statistical methods

8.1.1 Poisson distribution

The counts of oocysts in fixed volumes of water are subject to stochastic variation, i.e. different samples from the same body of water contain different numbers of oocysts. To compare visually the degree of variation between counts in raw and treated water the log-Normal distribution has been assumed in Figure 7.1 and Figure 7.2. This is not completely satisfactory for a number of reasons:-

1. The log-Normal distribution is a continuous distribution whereas the count data is discrete.

2. Zero counts found in samples cannot be adequately handled by the log-Normal since logarithm of zero is minus infinity. When estimating the parameters, whether by graphical methods or by method of maximum likelihood, the 'censoring' of a possibly large proportion of the data is unsatisfactory.

A more natural theoretical statistical distribution for discrete count data is the Poisson distribution. Here zero counts are perfectly acceptable. The chief assumptions required for the Poisson distribution are that the average rate of occurrence is fixed and the chance of occurrence is small enough to ignore the possibility of two or more simultaneous occurrences. The Poisson distribution has been found satisfactory in many practical cases, such as the number of faults in each square metre in sheets of glass, the number of radio-active particles detected by a Geiger counter in each second and the number of hits by V2 rockets in each quarter square mile during World War II.

The formula for the Poisson distribution is:

$$P(r) = \lambda^r e^{-\lambda} / r!$$

where $P(r)$ is the probability that r organisms will be observed and λ

is the Poisson parameter. The distribution has mean:

$$E(r) = \lambda$$

and variance:

$$Var(r) = \lambda.$$

An important property of the Poisson distribution is that it has only a single parameter, λ, which is equal to the mean density. However, the variance of the Poisson distribution is equal to the mean and hence is fixed once the mean is defined..

This inflexibility in the variance can be a disadvantage in many real life situations since the variance is often greater than the mean. This is known as "over-dispersion". It can occur where there the underlying mean density varies from place to place within the environment. This phenomenon is sometimes known as clustering.

When over-dispersion occurs, it is not valid to ignore it and continue with an analysis based on the Poisson. Two alternative methods of analysis can be used. Either another distribution, such as the Negative Binomial distribution, can be invoked, or the Poisson analysis can be modified to allow for the excess variation. Both methods were considered in this project.

8.1.2 Negative Binomial distribution

Suppose that the variation in the counts of an organism has a Poisson distribution with mean density λ locally, but that λ varies over a wider region according to the Gamma distribution. (The Gamma distribution is a continuous distribution, which takes only positive values and is skewed to the right.) Then it can be shown that the distribution of the organism over the wider area has a Negative Binomial distribution.

The Negative Binomial distribution can be expressed in a number of ways, but always has two parameters. We shall use the formula from Johnson and Kotz (1969):

$$P(r) = \binom{r+k-1}{k-1}\left(\frac{\mu}{\mu+k}\right)^r\left(\frac{k}{\mu+k}\right)^k$$

with $E(r) = \mu$ and $Var(r) = \mu+\mu^2/k = \mu(1+\mu/k)$

Thus μ is the location parameter and k is a scale parameter. It is clear that $Var(r) > E(r)$ provided $k>0$. The maximum likelihood estimate of μ for fixed k is simply the mean of the samples. k can be estimated using the method of moments by equating the sample variance to $\mu(1+\mu/k)$ and substituting the sample mean for μ, but this usually gives a relatively poor estimate. The maximum likelihood estimation of k involves an iterative procedure, but this can be performed with the statistical package GENSTAT, which also gives standard errors for the estimates.

Other distributions considered included Neyman's Type A (Neyman 1939) and Polya-Aeppli (Anscombe 1950), but these seemed less appropriate than the Negative Binomial since they involve recognisable clusters. In Neyman's Type A the number of clusters is distributed as Poisson (λ_1) and the number within a cluster is distributed as Poisson (λ_2). In the Polya-Aeppli the number of clusters is distributed as Poisson (λ) and the number within a cluster is distributed as Geometric(τ).

8.1.3 Over-dispersed Poisson and GLM

Where it is necessary to relate a response variable, such as the oocyst count, to explanatory variables, such as treatment parameters, the standard statistical procedure is to employ Generalised Linear Modelling (GLM). Where the distribution of the response variable is Poisson, the form of GLM generally used is called the log-linear model. Where different volumes are used, as in this exercise, it is possible to allow for this in the model by using the log of the volume as an offset.

However, where the response variable is Negative Binomial, constraints on the methodology restrict the use of GLM to cases where k is fixed and known. This is a severe drawback since generally there is no prior knowledge of the value of k.

In order to be able to perform a GLM analysis, a tractable solution is to fit a Poisson model but to allow the variance to exceed the mean. This over-dispersed Poisson model was fitted using GENSTAT, and the over-dispersion was taken into account in the estimation of standard errors of the estimates.

8.2 Results

8.2.1 Negative binomial distribution

Negative binomial distributions were fitted to the data from each of the four sampling experiments (3,5,6,7), separately for each sample (volume collected = 100 ml and 10 ml samples in the raw and 0.9 ml and 0.1 ml in the filtered water). The fitting was performed using the DISTRIBUTION directive in GENSTAT 5.3.1. Unfortunately, GENSTAT failed to converge in three cases for reasons not clear at present. The fitted cumulative distribution curves comparing observed and fitted values are illustrated for raw (Figure 8.1) and treated waters (Figure 8.2 and Figure 8.3) in experiment 5 and raw (Figure 8.4) and treated water (Figure 8.5) in experiment 6. The estimated parameters with standard errors are shown in Table 8.1. Two treated water samples with extreme high counts (488 and 359 per 100-ml) were excluded from the analyses for experiments 5 and 6 because they could not be fitted with the negative binomial distribution.

Figure 8.1 **Observed counts and cumulative negative binomial frequency distribution for spore counts in raw water samples (Experiment 5, 0.1 ml volume)**

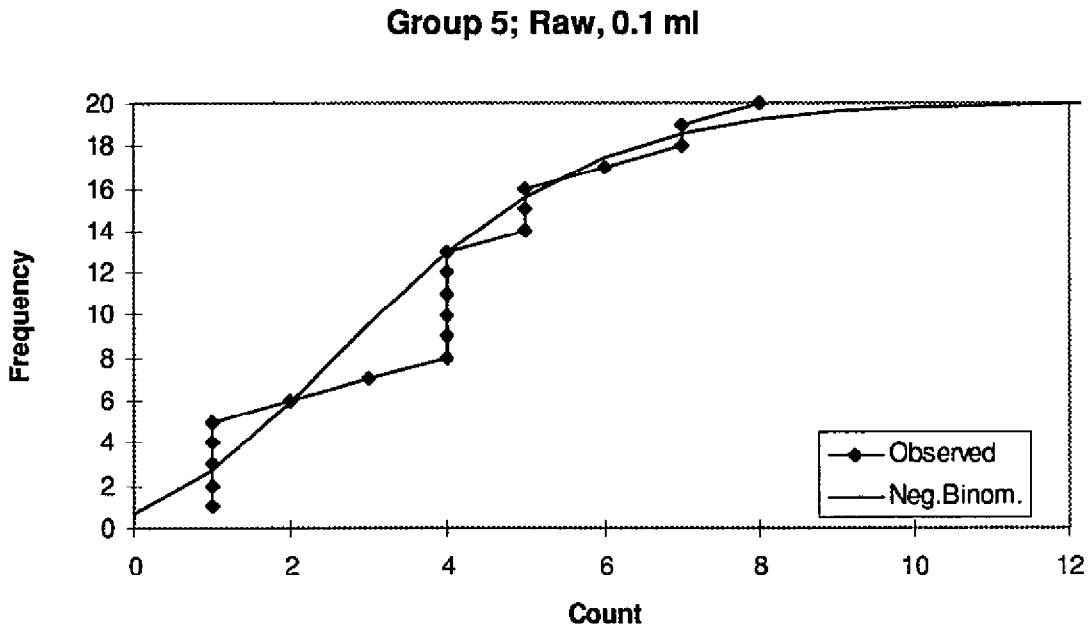


Figure 8.2 **Observed counts and cumulative negative binomial frequency distribution for spore counts in treated water samples (Experiment 5, 100 ml volume). Sample with 488 counts/100 ml excluded.**

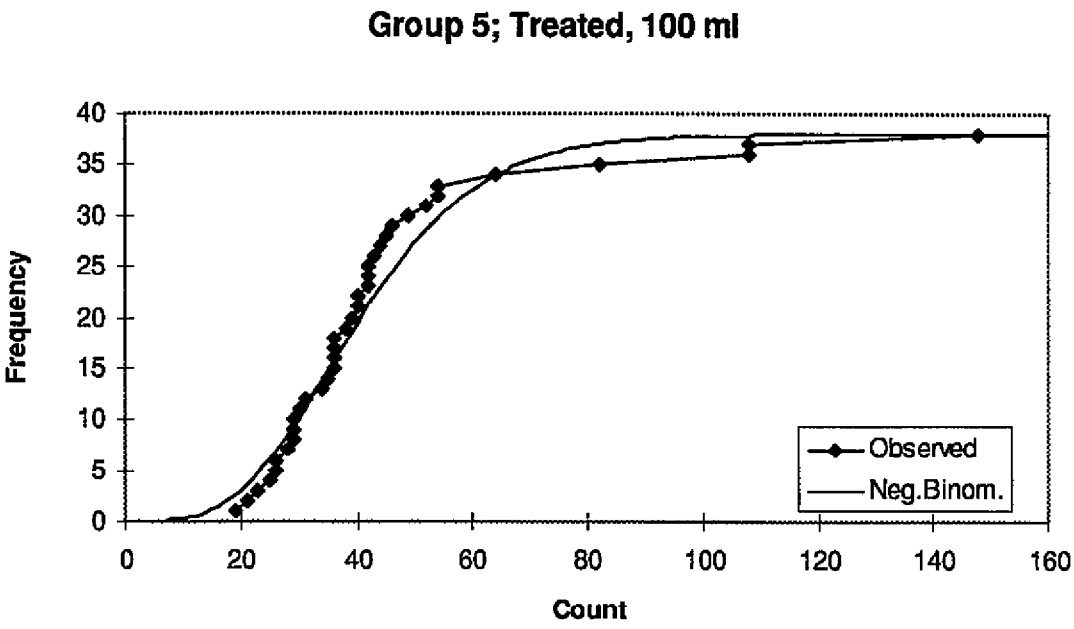


Figure 8.3 **Observed counts and cumulative negative binomial frequency distribution for spore counts in treated water samples (Experiment 5, 10 ml volume)**

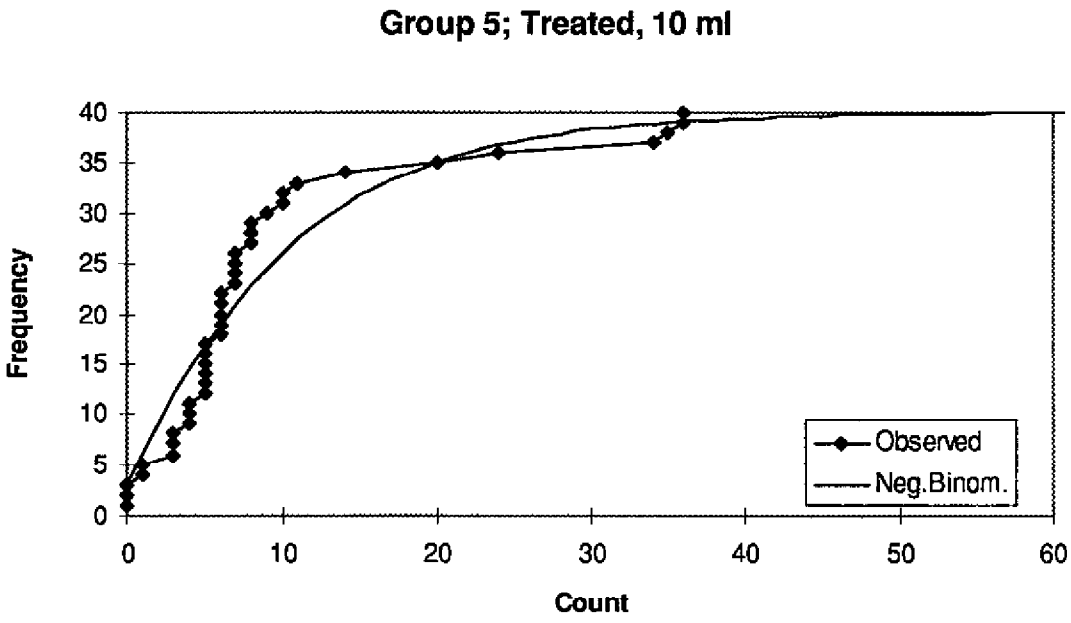


Figure 8.4 **Observed counts and cumulative negative binomial frequency distribution for spore counts in raw water samples (Experiment 6, 0.9 ml volume)**

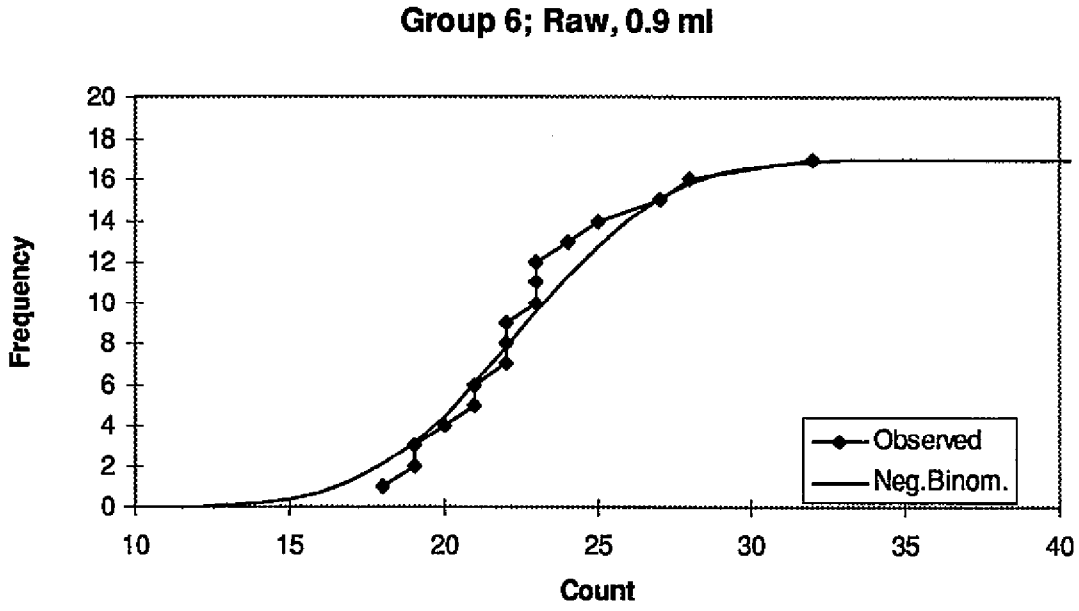
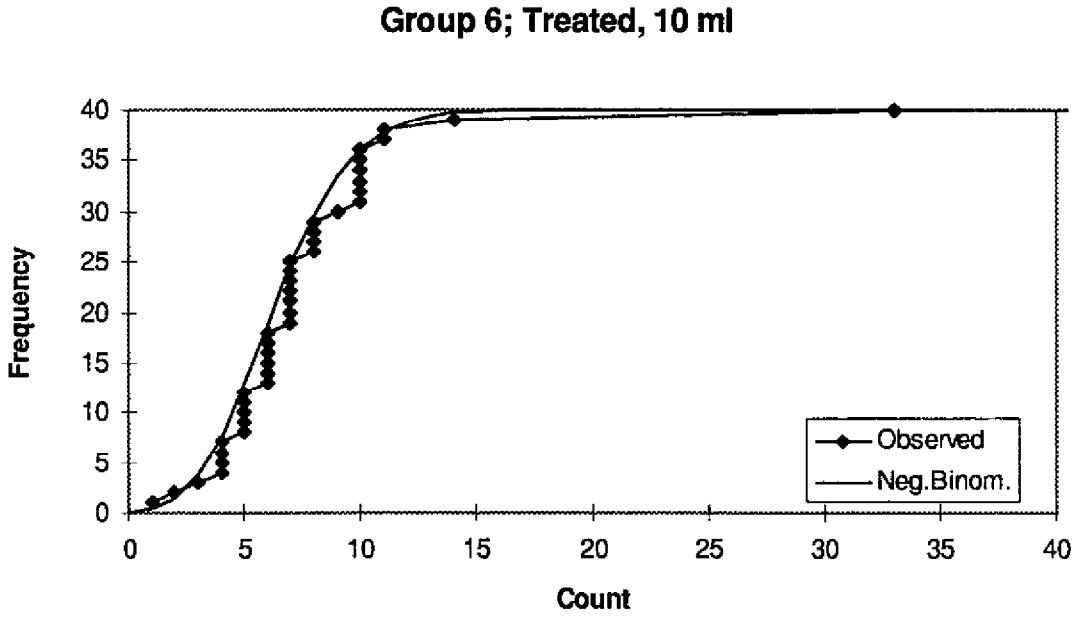


Figure 8.5 **Observed counts and cumulative negative binomial frequency distribution for spore counts in raw water samples (Experiment 6, 10 ml volume)**



The cumulative frequency curves indicate that the Negative Binomial distribution provides a good fit to the data. None of the fits were significantly poor when tested using the chi-square goodness-of-fit test (although with the small samples involved this is not a very powerful test).

Table 8.1 shows comparable means for the same combination of treatment and volume but considerable variation in the estimates for k. Negative estimates for k indicate that the variance of the data was less than the mean ('under-dispersed). The relatively large value of the standard errors for k reflect the sensitivity of k to small differences in the spread of observations about the mean.

Table 8.1 Fitted Negative Binomial parameters for spore data.

Experiment	Water type	Volume (ml)	No. of samples	Mean $\hat{\mu}$	s.e.	$\hat{k}$	s.e	θ
3	Filtered	10	33	10.0	0.96	5.5	2.7	2.81
3	Raw	0.9	8	24.7	0.73	-29.8	3.0	0.17
3	Raw	0.1	19	3.6	0.55	6.8	6.6	1.53
5	Filtered	100	39	45.0	3.44	5.0	1.2	10.03
5[1]	Filtered	100	38	42.2	2.84	7.0	1.8	7.07
5	Filtered	10	40	9.8	1.56	1.2	0.3	9.42
5	Raw	0.9	18	27.4	1.81	24.6	16.6	2.11
5	Raw	0.1	20	3.9	0.51	11.7	17.1	1.33
6	Filtered	100	39	68.3	1.81	83.3	42.2	1.82
6[2]	Filtered	100	38	68.4	1.86	75.4	36.8	1.91
6	Filtered	10	40	6.9	0.43	140.6	701.2	1.05
6	Raw	0.9	17	22.9	0.92	-59.8	32.2	0.62
6	Raw	0.1	20	[3]				
7	Filtered	10	32	[3]				
7	Raw	0.9	17	43.9	1.95	95.5	111.0	1.46
7	Raw	0.1	18	[3]				

[1]Excluding sample with 488 counts
[2]Excluding sample with 359 counts
[3]Failed to converge.

The variation between counts in 150 ml subsamples within the 100 litre treated water samples is accommodated by the fitted negative binomial distributions. Good fits are

apparent even when most of the samples with higher counts are included. Thus in Experiment 5, a negative binomial model could be fitted accommodating the four samples with over 30 spores per 10-ml despite most samples containing less than 10 spores (Figure 8.3). Similarly in Exercise 6 a negative binomial model could be fitted with one subsample containing 35 spores per 10-ml, despite most samples containing less than 15 spores (Figure 8.5). Similarly, in experiment 5, three 100-ml subsamples contained 100 to 160 spores despite most samples from that 100 litre volume of treated water containing fewer than 50 spores. Including those three samples, the curve fitted the rest of the data well (Figure 8.2). However, the fit was poor without excluding the high count sample with 488 spores per 100-ml. That point is therefore omitted in the fitted curve in Figure 8.2. Similarly it was necessary to exclude a 100-ml sample with 359 spores from the treated water in Exercise 6 to give a good fit to the rest of the data. It is concluded that the variations between counts in the majority of 10 or 100-ml subsamples within the 100 litre volumes of treated water are accommodated by the fitted Negative Binomial distributions but there is problem with rare occurrences of subsamples with extreme high counts. It was not possible to infer whether this problem was specific to this investigation or whether it occurs more generally.

The degree of variation between 150-ml subsamples from the raw waters was considerably lower (Figure 8.1 and Figure 8.4).

8.2.2 Log-linear Model

The log-linear model with a Poisson error structure has the following useful and important properties:

1. it enables a common model to be fitted to data observed in different volumes, in both raw and final water samples collected on different days;

2. this model provides estimates of the reduction in counts due to treatment; and

3. a dispersion parameter is estimated which represents the amount of over-dispersion compared with the Poisson distribution.

Two slightly different models were used. In the offset model, different sample volumes were accounted for by using the log of the sample volume as an offset. In the factor model, the volume enters the model as a categorical variable (known as a 'factor' in GENSTAT). Thus, the offset model is

$$\ln \mu_i = \ln v_i + \beta x_i$$

and the factor model is

$$\ln \mu_i = V_j \delta_{ij} + \beta x_i$$

where μ_i is the expected count in cell i,
v_i is the sample volume for cell i

x_i is the value of the explanatory variable in cell i,
β is a fitted coefficient for x_i
V_j is a fitted parameter for volume v_j
δ_{ij} is a dummy variable taking the value 1 if cell i had sample volume v_j and 0 otherwise

The explanatory variable was the water type, i.e. a factor taking the value 0 for Raw water and 1 for Filtered water.

The log-linear model was applied to each sampling exercise separately. Full results are shown in the Appendix, but estimates of the dispersion parameters are shown in Table 8.2. From this, it is apparent that exercises 3, 6, and 7 possessed similar dispersion parameters but that exercise 5 was much more variable. This reflected the greater variation in counts between subsamples from the 100 litre treated volume in Exercise 5.

Table 8.2 Estimates of dispersion parameters from log-linear model

Exercise	Offset model	Factor model
3	1.82	1.80
5	8.55	7.41
6	1.68	1.68
7	1.37	1.28

For the next stage of the analysis, data from exercises 3, 6 and 7 was combined and the offset model was fitted to the combined data. The dispersion index was estimated to be 2.19 and the fitted model for cell i:

$$\ln \mu_i = 3.23 - 3.17\omega_i - 0.39\gamma_{6i} + 0.87\gamma_{7i}$$

where ω and γ are dummy variables taking the values 0 or 1:
ω_i takes the value 0 for raw water and 1 for filtered water
g_{6i} takes the value 1 for exercise 6, 0 otherwise.
g_{7i} takes the value 1 for exercise 7, 0 otherwise.

(All four fitted parameters had standard errors between 0.05 and 0.07.)

Thus, water treatment effects an reduction of exp(3.17), i.e. a 24-fold reduction.

8.3 Discussion

The spatial variation in spore counts, in the volumes of water imbibed daily by consumers, within 100 litre volumes of drinking water can be modelled with the negative binomial distribution. However, a few samples contained very high counts which could not be accommodated by the negative binomial distribution. Although rare events those high count samples could be of importance in non-epidemic levels of waterborne cryptosporidiosis. To characterise them more fully, further sampling experiments are required.

It is clear from visual inspection of Figure 7.1 and Figure 7.2 that alum coagulation followed by rapid gravity filtration in drinking water treatment not only remove a considerable proportion of aerobic spores but also promote clustering. The degree of removal as determined by the log-linear model was 24-fold (i.e. 95.8% removal).

9. MODELLING THE VARIABILITY OF EXPOSURES TO TOTAL HETEROTROPHIC BACTERIA THROUGH DRINKING WATER

Densities of total heterotrophic bacteria (2 day, 37°C) collected over the period of a year from a drinking water supply zone in the UK are described by a log-normal distribution (Figure 9.1). Tap water consumptions by children and adults in the US has also been shown to be log-normally distributed (Roseberry and Burmaster 1992). For risk assessment modelling, however, it is the distribution of pathogen exposures through drinking water which is important. This would accommodate both the variation in pathogen density and tap water consumption. Exposures to bacterial doses through drinking water, being the product of bacterial density and drinking water consumption, may be calculated as:-

log (Bacterial Exposure) = log (Bacterial density) + log (Tap water consumption).

If ($\sigma_{\text{bacterial density}} = \sigma_{\text{tap water consumption}}$,) then the statistical distribution for bacterial exposure through drinking water could be modelled as the sum of the log-normal distributions for bacterial density and tap water consumption according to:-

Equation 1 $$N_{\text{exposure}}(\mu, \sigma^2) = N(\mu_{\text{bacterial density}}, \sigma_{\text{bacterial density}}^2) + N(\mu_{\text{tap water consumption}}, \sigma_{\text{tap water consumption}}^2)$$

where μ and σ are the mean and standard deviation, respectively, of the logarithms of bacterial densities and water consumptions and N represents the normal distribution. This may be rewritten as:-

Equation 2:- $$N_{\text{exposure}}(\mu, \sigma^2) = N(\mu_{\text{bacterial density}} + \mu_{\text{tap water consumption}}, \sigma_{\text{bacterial density}}^2 + \sigma_{\text{tap water consumption}}^2)$$

and suggests that bacterial exposures across the population through drinking water are also log-normally distributed. To test this, plate count exposures through drinking water were simulated using the GENSTAT statistical package (GENSTAT 5 Committee) and compared with the log-normal distribution for exposures calculated according to Equation 2 in Table 9.1. For the purposes of the exercise, μ and σ defining the log-normal distribution for plate count densities in Figure 9.1 and defining the log-normal distribution for tap water consumption in the 20 to 65 year age group (Roseberry and Burmaster, 1992) were used. To simulate exposures, 10,000 plate count and tap water consumption values were generated with GENSTAT using the parameters μ and σ defining each distribution and added together. The simulated (logarithm) exposures were shown to be normal giving a straight line on a normal probability plot. Furthermore, from Table 9.1 it is apparent that μ and σ calculated from the simulated (logarithm) exposures were very similar to those calculated by summing the two normal distributions according to Equation 2.

Table 9.1 Statistics for log-normal distributions to calculate daily exposure to plate count bacteria through drinking water

Distribution	Units	Mean of $\log_{10}$ transformed) data (μ	Standard Deviation of $\log_{10}$ transformed data (σ)	Percentiles		
				2.5%	50%	97.5%
Bacterial density	plate count/ml	0.7089[a]	0.768[b]	0.16	5.11	163.7
Tap water consumption	ml/person /day	3.0500[c]	0.2123[c]	430	1,122	2,924
Bacterial exposure calculated as the sum of the above distributions	plate count /person/day	3.7589[d]	0.7968[e]	157	5,739	209,231
Bacterial exposure simulated by GENSTAT	plate count /person/day	3.769[f]	0.7987[f]	159[g]	5,825[g]	223,100[g]

[a]calculated as intercept of best fit line with 50 percentile (Figure 9.1)
[b]calculated as slope of best fit line (Figure 9.1)
[c]taken from Roseberry and Burmaster (1992) and transformed from base e to base 10
[d]calculated as $\mu_{\text{bacterial density}} + \mu_{\text{tap water consumption}}$
[e]calculated as $\sqrt{(\sigma_{\text{bacterial density}}^{2} + \sigma_{\text{tap water consumption}}^{2})}$
[f]calculated from exposures simulated using GENSTAT
[g]exposures simulated by GENSTAT

Table 9.1 also shows the different percentiles for the three log-normal distributions calculated from the formulae:-

2.5 pecentile = $10^{(\mu - 1.96\sigma)}$.

50 pecentile (geometric mean) = $10^{(\mu)}$.

97.5 pecentile = $10^{(\mu + 1.96\sigma)}$.

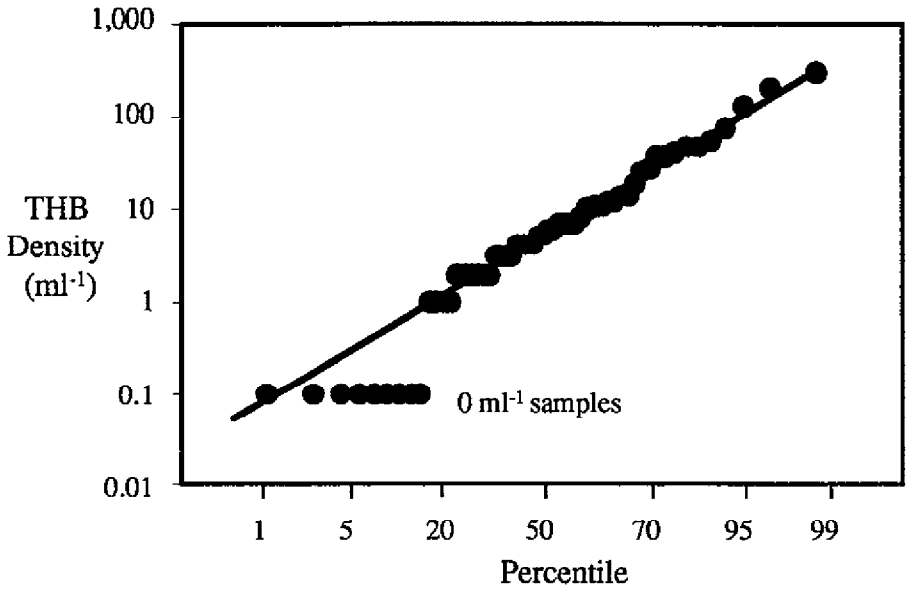
Corresponding percentiles from the simulated exposure values agree well with those calculated from the log-normal distribution for exposure (Equation 2). While the least exposed 2.5% of tap water drinkers in that UK supply zone consume <157 plate count

bacteria per day, the most exposed 2.5% in that same zone consume >209,000 plate count bacteria per day, i.e. over 1,300-fold more.

9.1 Discussion

Simulations have been performed in this section to predict the variability in exposures to daily doses of total heterotrophic bacteria through drinking water. These simulations take into account not only the variability in volumes of drinking water consumption across the population but also the variability in densities of bacteria across a UK drinking water supply zone during the year. The variation in daily exposures was large, calculated at over 1,300-fold across 95% of the population.

Figure 9.1 **Log-normal distribution for densities of total heterotrophic bacteria collected from a drinking water supply zone in the UK in 1990.**



10. PREDICTING *CRYPTOSPORIDIUM* EXPOSURES THROUGH DRINKING WATER

The approach taken here is to use the data for *Cryptosporidium* densities in the raw water and the treatment removal efficiency data to predict the oocyst densities in drinking water.

10.1 Applying log-normal distributions to microbiological risk assessment for *Cryptosporidium* in drinking water

This is dependent on knowing the parameters (μ and σ) defining the log-normal distribution for pathogen densities in the drinking water supply. There is very little information on pathogen densities in the drinking water supply because the majority of large volume samples (up to 1,000 litres) are negative (Payment *et al.* 1985; LeChevallier *et al.* 1991). This may be because pathogens are generally dilute across the whole supply or because they are highly clustered and hence not encountered without intensive sampling. A further problem is that the parameters μ and σ will vary depending on the efficiency of pathogen removal by drinking water treatment and also on the microbial loading of the raw water.

A possible solution is to estimate the log-normal distribution of pathogens in the drinking water supply from:-

- the log-normal distribution of pathogens in the raw water
- the log-removal efficiency for each pathogen by a particular drinking water treatment process.

This approach would enable the risks on health of:-

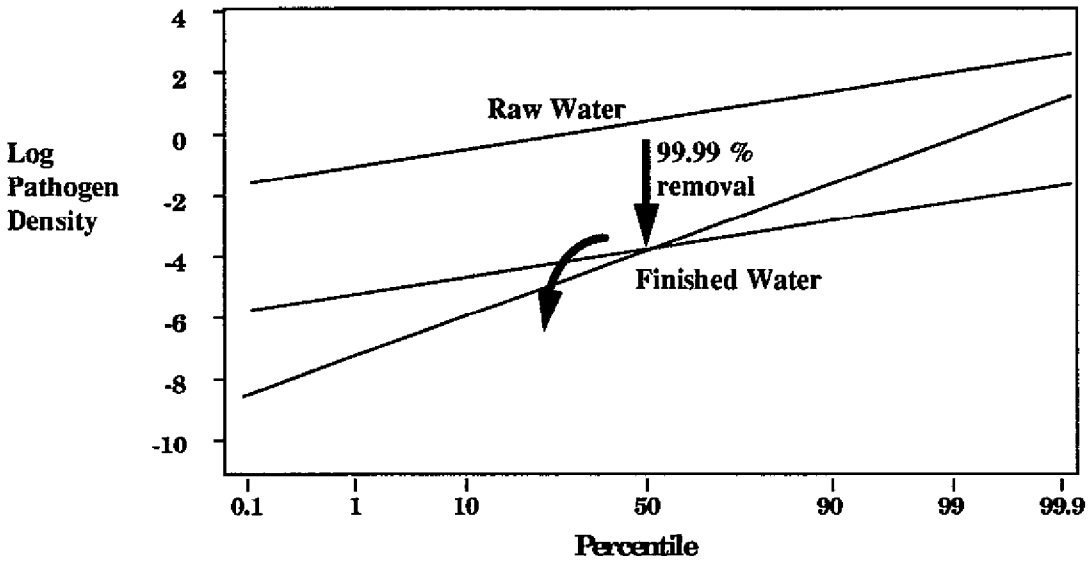
1. different drinking water treatment processes and
2. increases in microbial loadings in the raw waters

to be modelled.

The approach is shown in Figure 10.1. Pathogen densities in a raw surface water are plotted as a log-normal distribution. The predicted distributions for pathogen densities in the drinking water supply after median pathogen densities in raw water are reduced by 99.99% are shown for two scenarios:-

1. pathogen clustering is not promoted by drinking water treatment
2. pathogen clustering is promoted by drinking water treatment

Figure 10.1 Predicting pathogen densities in the drinking water supply assuming that median densities in the raw water are reduced by 99.99%



10.1.1 Modelling exposures to *Cryptosporidium* oocysts through drinking water assuming pathogen clustering is not promoted by drinking water treatment.

The approach to predicting pathogen exposures is performed here using *Cryptosporidium* oocyst data for surface water systems presented in Table 5.1 and Table 5.2. For two rivers (A, with agricultural drainage; B, with rural domestic drainage) all oocysts densities recorded were above the limits of detection and the parameters μ and σ were measured from the best fit lines describing log-normal distributions for oocyst densities in Figure 1 of Hutton *et al.*(1995). These parameters are presented for Rivers A and B in Table 5.1. Variation in oocyst densities were much higher in two other surface water systems(Table 5.1) for which 55% to 90% of densities were below the limits of detection. To assign values to those densities would have required analysis of much larger sample volumes. Therefore information on the statistical distribution is incomplete. For the purposes of continuing this exercise, the assumption is made that the best fit lines for each distribution could be extrapolated back to define the log-normal distributions (see Figure 5.1). Log-normal distributions for oocyst densities in the drinking water supply after treatment of water from each of the five surface waters were modelled by subtracting 2.0 from μ (Table 10.1). It was assumed that treatment had the same effect on each density and therefore σ was not changed. This is equivalent to a treatment process which removes 99% of oocysts from each and every density.

Table 10.1 Parameters defining log-normal distributions for *Cryptosporidium* oocyst densities in five surface waters and predicted for drinking water supplies after a treatment process which removed 99% of oocysts.

Water Supply	Estimated Surface Water (log_{10} count/litres)		Predicted Drinking Water (log_{10} count/litres)	
	μ	σ	μ	σ
River A (agricultural drainage)	0.969	0.575	-1.031	0.757
River B (rural domestic drainage)	0.212	0.424	-1.788	0.424
Farmoor Reservoir[a].	-2.5	1.575	-4.5	1.575
River Tillingbourne[a]	-2.0	1.88	-4.0	1.88
Protected Watershed (*Giardia*)[a]	-1.575	1.00	-3.575	1.00

[a]To estimate μand σ for pathogen densities in raw source water, best fit lines defining log-normal distributions were extrapolated because 55 - 90% of samples were below limits of detection (Hutton *et al.* 1995).

Exposures to oocysts through drinking water were simulated using the log-normal distribution for oocysts densities predicted for each drinking water supply (Table 10.1) and the log-normal distribution for tap water consumption in the 20 to 65 year age group (Roseberry and Burmaster 1992) using the statistical package GENSTAT. Simulated percentiles for oocyst exposures are presented in Table 10.2. It is apparent that the possible variation in daily oocyst exposures is large. To give some idea of exposure variation, the ratios of the 97.5% to the 2.5% percentile (which is equivalent to the variation in oocyst exposures across 95% of consumers) are presented (Table 10.2). The model predicts that oocyst exposures across 95% of consumers through drinking water vary 70 and 250-fold for drinking water systems supplied by the two rivers with complete oocyst density data. For systems with larger oocyst density variation, exposures across the population varies much more; perhaps in the order of 10^4 to 10^6-fold (Table 10.2).

Table 10.2 Variations in simulated exposures to *Cryptosporidium* oocysts across the population of tap water consumers after 99% oocyst removal from the raw water. Log-normal distribution for pathogen density in drinking water defined by parameters in Table 10.1. Log-normal distribution for tap water consumption for 20 - 65 year age group(Roseberry and Burmaster 1992) was used.

Location	Percentiles for exposure (oocysts/ person/day)			Variation across 95%
	2.5%	50%	97.5%	of consumers
River A (agricultural drainage)	0.0067	0.1051	1.73	255
River B (rural domestic drainage)	0.00220	0.0185	0.159	72
Farmoor Reservoir	0.000000028	0.000036	0.0531	1.9×10^6
River Tillingbourne	0.0000000228	0.000117	0.6712	2.9×10^7
Protected Watershed (*Giardia*)	0.0000030	0.000305	0.0328	1.1×10^4

It is apparent from the percentiles simulated for exposure (Table 10.2), that the major proportion (97.5%) of consumers are predicted to be exposed to very low doses (<1 pathogen/person/day) of *Cryptosporidium* oocysts (or *Giardia* cysts). In the case of the River Tillingbourne-treated water (with the largest variation in oocyst density, Table 10.1) 0.56% of the 10,000 simulated exposures exceeded 10 *Cryptosporidium* oocysts/person/day and 0.10% of simulated exposure exceeded 100 oocysts/person/day.

The largest simulated exposure was 268 oocysts/person/day. The model therefore predicted that a small proportion of consumers to be exposed to high doses of pathogen through drinking water. In the case of the drinking water supply using River A as source, 0.08% of 10,000 simulated exposures exceeded 10 oocysts/person/day.

Discussion

This approach only takes into account the variability in pathogen densities between large (100 litre or 1,000 litre) volume samples. It does not take into account that finding in Sections 7 and 8 that drinking water treatment promotes spatial association of micro-organisms. The assumption is made that oocysts are Poisson distributed in 100 litre volume drinking water samples which is incorrect. The variability in exposures to oocysts through drinking water appears to be large reflecting the large variability in oocyst densities between large volume samples. However, even with only a 99% removal of oocysts by treatment, the exposures are so small (less than 2 oocysts per person per day in 97.5% of the population, Table 10.2) that most consumers are not likely to be exposed to high doses through drinking water. However, the large variability in oocyst densities between large volume raw water samples is manifest in a small proportion (0.1%) of the population being exposed to high doses (>100 oocysts/person/day) after a 99% oocyst removal by drinking water treatment. The simulations also do not allow for the fact that in the UK only 10% of tap water is consumed unboiled (Table 6.3).

10.1.2 Modelling exposures to *Cryptosporidium* oocysts through drinking water assuming pathogen clustering is promoted by drinking water treatment.

The simulations performed above do not take into account the findings from Sections 7 and 8 that micro-organisms are not randomly dispersed through 100 litre volume treated water samples but are spatially associated to some degree. To accommodate these findings the distribution of oocysts in each 100 litre volume should be modelled using the negative binomial distribution (e.g. with parameters Table 8.1). Furthermore, account needs to be taken of the rare occurrence of samples with much higher spore counts than accommodated by the negative binomial distribution (e.g. Table 7.4). Accounting for this would mean some consumers are exposed to higher doses.

11. DOSE-RESPONSE CURVE IN CRYPTOSPORIDIUM RISK ASSESSMENT MODEL

The dose-response curve is central to current microbiological risk assessment models (Rose and Gerba, 1991). Recently dose response data have been published for *Cryptosporidium* in healthy adult human volunteer studies (DuPont *et al.* 1995). The dose response curve modelled through those data with parameters presented by Rose *et al.* (1995) is shown in Figure 11.1 and would be used in the *Cryptosporidium* risk model for healthy adults in the UK.

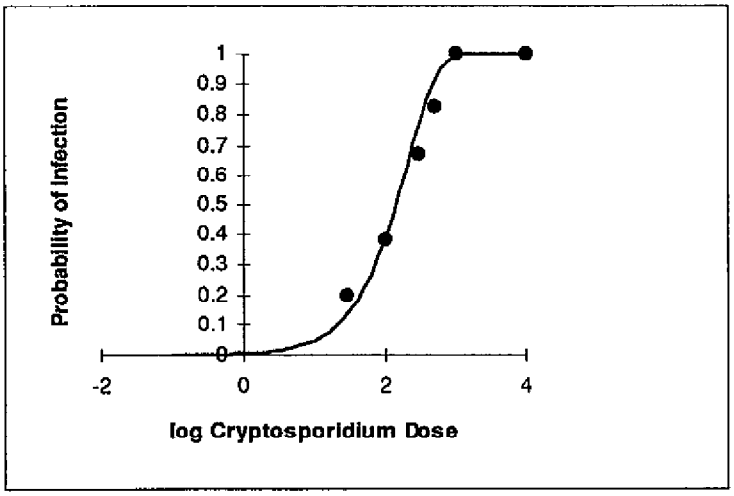


Figure 11.1 Dose-Response curve for *Cryptosporidium* in healthy adults. Data from DuPont *et al* (1995). Single-hit exponential model ($P = 1 - e^{-(0.0047N)}$)

12. CONCLUSIONS

Microbiological risk assessment models have been used to define targets for the degree of pathogen removal by drinking water treatment (Surface Water Treatment Rule) and to identify the maximum concentrations of pathogens in drinking water (Maximum Contaminant Levels) to ensure acceptable risk from waterborne pathogens to consumers.

Current risk assessment models developed for *Cryptosporidium*, *Giardia* and enteric viruses in drinking water use the Beta-Poisson or negative exponential form of dose-response curves. That approach relies on the assumption that pathogen organisms are randomly dispersed within 100 litre or 1,000 litre volumes. Available evidence in the literature shows this assumption is not appropriate in drinking water. Experiments performed in this contract have demonstrated that aerobic spores, which may serve as surrogates for *Cryptosporidium* oocysts, are not randomly dispersed within 100 litre volumes of drinking water but are spatially associated to some degree. This spatial association is manifest in some 100-ml subsamples containing considerably higher spore counts than others. It has been possible to accommodate some of the spatial variation by fitting the negative binomial statistical distribution. However a small number of 100-ml samples contained very high counts which could not be accommodated by the negative binomial distribution. Although rare events those high count samples could be of importance in non-epidemic levels of waterborne cryptosporidiosis and need to be included in risk assessment models.

The main components to develop a model to predict the risks of infection through consumption of drinking water in the UK have been assembled in this contract. The components include:-

- log-normal distribution for oocyst densities in UK source waters
- negative binomial models for distribution of oocysts within 100 litre treated water volumes at the resolution of 100-ml
- log-normal distribution for unboiled tap water consumption in the UK
- dose-response curve relating oocyst exposure to the probability of infection

Before a *Cryptosporidium* risk assessment model can be constructed further statistical consideration is needed to:-

1. relate negative binomial distributions for counts in raw water samples to counts in treated water samples, that is, define the effect of treatment in terms of the parameters describing the negative binomial distribution.
2. accommodate the few samples with very high counts not accounted for by the negative binomial distribution.

From experiments with aerobic spores reported here, the data necessary to achieve the first of these two tasks are available. However, to characterise the rare occurrences of very high counts, much larger numbers of samples will be required.

The advantage of developing risk assessment models along these lines is that it would enable the effect of clustering on the risk of waterborne cryptosporidiosis to be defined. In the experiments performed here to assess variability between samples in spore counts a mild detergent, Tween, was used to disperse spores attached to particulates. Such treatment may not be fully effective at breaking up and releasing spores from biofilm and, in particular, metal oxides from chemical coagulation. Therefore, it is possible that through oocysts bound in high numbers to Alum particles breaking through into distribution, some consumers may be exposed to higher doses than predicted by models developed in this work.

Overall, the data and methodology to develop a model for assessing risks from waterborne cryptosporidiosis have been established in this contract.

Simulations have been performed in this contract to predict the variability in exposures to total heterotrophic bacteria through drinking water. These simulations take into account not only the variability in volumes of drinking water consumption across the population but also the variability in densities of bacteria across a UK drinking water supply zone during the year. The variation in daily exposures was large, calculated at over 1,300-fold across 95% of the population.

Simulations have also been performed to predict the variability in exposures to *Cryptosporidium* oocysts based on log-normal distributions for densities in the raw water and consumption of tap water. Even with only a 99% removal of oocysts through drinking water treatment, the exposures are so small that most consumers are not likely to be exposed to high doses. Accounting for the increase in oocyst clustering as judged by spore experiments performed in this work would, however, mean some consumers are exposed to higher doses.

Patterns of drinking water consumption in the UK have been reviewed. It would appear that only 10% of tap water is consumed unboiled. The proportion of tap water consumed unboiled is highest among infants but lowest in the elderly.

13. GLOSSARY

Exp, exponential function

Gamma distribution, Skewed statistical distribution for continuous data limited to positive values

Generalised linear models, statistical models that are a natural extension of the classical linear model and that share a common method for computing parameter estimates

GENSTAT, Statistical software package

ID_{50}, (Infectious dose) number of pathogens to infect 50% of population

Immunosuppressed, persons or animals whose immune system has temporarily been suppressed by means of drugs or irridiation

Isoelectric point, conditions under which molecule has neutral charge

Log normal distribution, Skewed statistical distribution for continuous data of positive value

Negative binomial distribution, statistical distribution for discrete counts

Neyman Type A, statistical distribution for discrete counts

PAC, poly aluminium chloride

Percentile, a partition value based on percentages; for example the 90 percentile is the value that separates a frequency distribution into the 90% of smallest values and the 10% of greatest values

Polya-Aeppli, a statistical distribution for discrete counts

Poisson distribution, statistical distribution for discrete counts

Gnotobiotic, disease-free

TTC, 2,3,5-triphenyl-tetrazolium chloride for detecting metabolically active cells

Zeta potential, potential at surface of shear

14. REFERENCES

Anscombe, F.J. (1950) Sampling theory of the negative-binomial and logarithmic series distributions. *Biometrika*, **37**, 358

Blewett, D.A., Wright, S.E., Casemore, D.P., Booth, N.E. and Jones, C.E.(1993). Infective dose size studies on *Cryptosporidium parvum* using gnotobiotic lambs. *Water Science and Technology*, **27**, 61-64

Carrington, E.G. and Ransome, M.E. (1994). *Factors influencing the survival of* Cryptosporidium *oocysts in the environment.* [Report FR 043] Marlow: Foundation for Water Research.

Carrington, E.G. and Smith, H.V. (1995). The occurrence of Cryptosporidium spp. oocysts in surface waters and factors influencing their levels, with particular reference to the United Kingdom. In: Betts, W.B., Casemore, D.P., Fricker, C.R., Smith, H.V. and Watkins, J., Editors. *Protozoan parasites and water.* Cambridge: Royal Society of Chemistry. 57-62.

Casemore, D.P. and Roberts, C. (1993) Guidelines for screening for *Cryptosporidium* in stools. Report of a joint working group. *Journal of Clinical Pathology*, **46**, 2-4.

Casemore, D.P. (1995) The problem with protozoan parasites. In: Betts, W.B., Casemore, D.P., Fricker, C.R., Smith, H.V. and Watkins, J., Editors. *Protozoan parasites and water.* Cambridge: Royal Society of Chemistry. 10-18.

Chambers, R.M., Sturdee, A.P., Bull, S.A. and Miller, A. (1995). Rodent reservoirs of *Cryptosporidium.* In: Betts, W.B., Casemore, D.P., Fricker, C.R., Smith, H.V. and Watkins, J., Editors. *Protozoan parasites and water.* Cambridge: Royal Society of Chemistry. 63-66

Christian, R.R. and Pipes, W.O. (1983) Frequency distribution of coliforms in water distribution systems. *Applied and Environmental Microbiology*, **45**, 603-609.

Communicable Disease Report (1995). AIDS and HIV-1 infection in the United Kingdom. *Communicable Disease Report*, **5**, 289-290

Department of the Environment and the Department of Health (1995). *Cryptosporidium in water supplies - Second report of the Group of Experts.* London:HMSO

Department of Health (1995). Advice to boil water to prevent cryptosporidiosis in immuno-compromised patients, CMO'

DuPont, H.L., Chappell, C.L., Sterling, C.R., Okhuysen, P.C., Rose, J.B. and Jakubowski, W. (1995). The infectivity of *Cryptosporidium parvum* in healthy volunteers. *New England Medical Journal*, **322**, 855-859.

Gergory, M.W. (1990). Epidemiology of cryptosporidiosis. In: *Cryptosporidium in water supplies - Report of the Group of Experts*. London: HMSO. 82-99.

Gregory, J., Ives, K.J., Scutt, J.E., and Makanjuola, D.B., (1991) Removal of Cryptosporidium oocysts by water treatment methods at laboratory scale. Final report on work carried out at University College London, March 1990 - March 1991.

Haas, C.N., Rose, J.B., Gerba, C. and Regli, S. (1993) Risk assessment of virus in drinking water. *Risk Analysis*, **13**, 545-552.

Hall, T., Pressdee, J. and Carrington, E. (1995) Removal of *Cryptosporidium* during water treatment. In *Protozoa Parasites and Water* ed. Betts, W.B., Casemore, D., Fricker, C., Smith, H. and Watkins, J. pp.192-195. The Royal Society of Chemistry.

Hopkin, S.M. and Ellis, J.C. (1980) Drinking Water Consumption in Great Britain - a survey of drinking habits with special reference to tap-water based beverages. WRc technical Report TR 137.

Hutton, P., Ashbolt, N., Vesey, G., Walker, J. and Ongerth, J. (1995) *Cryptosporidium* and *Giardia* in the aquatic environment of Sydney, Australia. In *Protozoa Parasites and Water* ed. Betts, W.B., Casemore, D., Fricker, C., Smith, H. and Watkins, J. pp.71-75. The Royal Society of Chemistry.

Johnson, N.L. and Kotz, S. (1969) *Discrete distributions*. Houghton Mifflin, Boston, Massachussets

Kemp, J.S., Wright, S.E. and Bukhari, Z. (1995). On farm detection of *Cryptosporidium parvum* in cattle, calves and environmental samples. In: Betts, W.B., Casemore, D.P., Fricker, C.R., Smith, H.V. and Watkins, J., Editors. *Protozoan parasites and water*. Cambridge: Royal Society of Chemistry. 154-157.

LeChevallier M.W., Norton, W.D. and Lee, R.G. (1991) *Giardia* and *Cryptosporidium spp.* in filtered drinking water supplies. *Applied and Environmental Microbiology*, **57**, 2617-2621.

LeChevallier, M.W. and Norton, W.D. (1995) Occurrence of *Giardia* and *Cryptosporidium* in raw and finished water. *Journal American Waterworks Association*, **87**; 9: 54-68.

Marshall, R., Holmes, E. and Stanwell-Smith, R. (1991). *Cryptosporidiosis surveillance 1990-1991*. [Report to the Water Research Centre]

McDonald, V., and Award-el-Kariem, F.M. (1995). Strain variation in *Cryptosporidium parvum* and evidence for distinctive human and animal strains. In: Betts, W.B., Casemore, D.P., Fricker, C.R., Smith, H.V. and Watkins, J., Editors. *Protozoan parasites and water*. Cambridge: Royal Society of Chemistry. 57-62.

MacKenzie, W.R., Hoxie, N.J., Proctor, M.E., Gradus, M.S., Blair, K.A., Peterson, D.E., Kazmierczak, J.J., Addiss, D.G., Fox, K.R., Rose, J.B. and Davis, J.P. (1994) A massive

outbreak in Milwaukee of *Cryptosporidium* infection transmitted through the public water supply. *New England Journal of Medicine*, **331**, 161-167.

Maul, A., El-Shaarawi, A.H. and Block, J.C. (1985) Heterotrophic bacteria in water distribution. I. Spatial and temporal variation. *The Science of the Total Environment*, **44**, 201-214.

Medema, G.J., Teunis, P.F.M., Gornik, V., Havelaar, A.H. and Exner, M. (1995) Estimation of the *Cryptosporidium* Infection Risk via Drinking Water. In *Protozoa Parasites and Water* ed. Betts, W.B., Casemore, D., Fricker, C., Smith, H. and Watkins, J. pp.53-56. The Royal Society of Chemistry.

MEL Research (1995) Drinking Water Consumption Study Draft Final Report to the Drinking Water Inspectorate.

Miller, R.A., Bronsdon, A. and Morton, W.R. (1990). Experimental cryptosporidiosis in a primate model. *Journal of Infectious Diseases*, **161**, 2071-2075.

National Cryptosporidium Survey Group (1992). A survey of *Cryptosporidium* oocysts in surface and groundwaters in the UK. *Journal of the Institute of Water and Environmental Management*, **6**, 697-703.

Neyman, J. (1939) On a new class of "contagious" distributions, applicable in entomology and bacteriology. *Ann. Math. Statist.*, **10**, 35

Payment, P., Trudel, M. and Plante, R. (1985) Elimination of viruses and indicator bacteria at each step of treatment during preparation of drinking water at seven water treatment plants. *Applied and Environmental Microbiology*, **49**, 1418-1418.

Pipes, W.O., Ward, P. and Ahn, S.H. (1977) Frequency distributions for coliform bacteria in water. *Journal American Waterworks Association*, **69**, 664-668.

Ransome, M.E., Whitmore, T.N. and Carrington, E.G. (1993) Effect of disinfectants on the viability of *Cryptosporidium parvum* oocysts. *Water Supply*, 11, 75-89.

Regli, S., Rose, J.B., Haas, C.N. and Gerba, C.P. (1991) Modelling the risk from *Giardia* and viruses in drinking water. *Journal American Waterworks Association*, **83**, 11: 76-84.

Robertson, L.J. (1995) Viability of *Giardia* cysts and *Cryptosporidium* oocysts. In: Betts, W.B., Casemore, D.P., Fricker, C.R., Smith, H.V. and Watkins, J., Editors. *Protozoan parasites and water*. Cambridge: Royal Society of Chemistry. 97-103.

Robertson, L.J., Campbell, A.L. and Smith, H.V. (1992). Survival of *Cryptosporidium parvum* oocysts under various environmental pressures. *Applied and Environmental Microbiology*, **58**, 3494-3500.

Rogers, J. and Keevil, C.W. (1995) Survival of *Cryptosporidium parvum* oocysts in biofilm and planktonic samples in a model system. In *Protozoa Parasites and Water* ed.

Betts, W.B., Casemore, D., Fricker, C., Smith, H. and Watkins, J. pp. 209-213. The Royal Society of Chemistry.

Rose, J.B. and Gerba, C.P. (1991) Use of risk assessment for development of microbial standards. *Water Science and Technology.*, **24**, 29-34.

Rose, J.B., Haas, C.N. and Regli, S. (1991a) Risk assessment and control of waterborne giardiasis. *American Journal of Public Health*, **81**, 709-713.

Rose, J.B., Gerba, C.P. and Jakubowski, W. (1991b) Survey of potable water supplies for *Cryptosporidium* and *Giardia. Environmental Science and Technology*, **25**, 1393-1400.

Rose, J.B., Lisle, J.T. and Haas, C.N. (1995) Risk Assessment Methods for *Cryptosporidium* and *Giardia* in contaminated water. In *Protozoa Parasites and Water* ed. Betts, W.B., Casemore, D., Fricker, C., Smith, H. and Watkins, J. pp.238-242. The Royal Society of Chemistry.

Roseberry, A. and Burmaster, D.E. (1992) Log-normal distributions for water intake by children and adults. *Risk Analysis*, **12**, 99-104.

Scott, C.A., Smith, H.V. and Gibbs, H.A. (1994). Excretion of *Cryptosporidium parvum* oocysts by a herd of beef suckler cows. *The Veterinary Record,* **134**, 170.

Smith, H.V., Sayer, P.D., Ngotho, J.M., Ndung'n, J.M., Ngure, R.M. and Grimason, A.M. (1993). Determination of the minimum infective dose of *Cryptosporidium parvum* oocysts in vervet monkeys. [Final Report to the Department of the Environment].

Smith, H.V., Sayer, P.D., Bukhari, Z., Ngotho, J.M., Ndung'n, J.M., Ngure, R.M. and Grimason, A.M. (1994). Determination of the IgG antibody levels to *Cryptosporidium parvum* oocyst antigens in vervet monkeys infected with various doses of human-derived *C. parvum* oocysts. [Addendum to Final Report to the Department of the Environment].

Smith, H.V. (1992) *Cryptosporidium* and water: a review. *Journal of the Institution of Water Engineers and Managers*, **6**, 443-450.

Sorvillo, F., Lieb,L.E.,Nahlen, B., Miller, J., Mascola, L. and Ash, L.R. (1994). Municipal drinking water and AIDS in Los Angeles County. *Epidemioogy and Infection*, **113**, 313-320.

Ward, R.L., Bernstein, D.I., Young, E.C., Sherwood, J.R., Knowlton, D.R. and Schiff, G.M.(1986) Human rotavirus studies in volunteers: Determination of infectious dose and serological response to infection. *Journal of Infectious Diseases*, **154**, 871-880.

15. APPENDIX

15.1 Sampling Exercise 3

15.1.1 Offset model

*** Summary of analysis ***

	d.f.	deviance	mean deviance	deviance ratio
Regression	1	1135.0	1134.953	622.17
Residual	58	105.8	1.824	
Total	59	1240.8	21.030	

Dispersion parameter is estimated to be 1.82

*** Estimates of regression coefficients ***

	estimate	s.e.	t(58)
Constant	3.3715	0.0829	40.67
Water filtered	-3.418	0.112	-30.41

15.1.2 Factor model

*** Summary of analysis ***

	d.f.	deviance	mean deviance	deviance ratio
Regression	2	225.3	112.670	62.77
Residual	57	102.3	1.795	
Total	59	327.7	5.554	

*** Estimates of regression coefficients ***

(Volume 10) = (Water filtered)

	estimate	s.e.	t(57)
Constant	1.275	0.162	7.85
Water filtered	0.981	0.179	5.48
Volume 0.9000	1.929	0.188	10.24
Volume 10	0	*	*

15.2 Sampling Exercise 5

15.2.1 Offset model

*** Summary of analysis ***

	d.f.	deviance	mean deviance	deviance ratio
Regression	1	3434.1	3434.135	401.72
Residual	114	974.5	8.549	
Total	115	4408.7	38.336	

*** Estimates of regression coefficients ***

	estimate	s.e.	t(114)
Constant	3.437	0.123	27.97
Water filtered	-4.138	0.139	-29.86

15.2.2 Factor model

*** Summary of analysis ***

	d.f.	deviance	mean deviance	deviance ratio
Regression	3	1544.2	514.734	69.38
Residual	112	830.9	7.419	
Total	115	2375.1	20.653	

*** Estimates of regression coefficients ***

(Volume 100) = (Water filtered) - (Volume 10)

	estimate	s.e.	t(112)
Constant	1.348	0.310	4.34
Water filtered	2.458	0.317	7.75
Volume 0.9000	1.954	0.334	5.85
Volume 10	-1.571	0.155	-10.11
Volume 100	0	*	*

15.3 Sampling Exercise 6

15.3.1 Offset model

*** Summary of analysis ***

	d.f.	deviance	mean deviance	deviance ratio
Regression	1	2224.0	2223.953	1326.25
Residual	113	189.5	1.677	
Total	114	2413.4	21.171	

*** Estimates of regression coefficients ***

	estimate	s.e.	t(113)
Constant	3.2246	0.0621	51.94
Water filtered	-3.5964	0.0666	-54.02

15.3.2 Factor model

*** Summary of analysis ***

	d.f.	deviance	mean deviance	deviance ratio
Regression	3	3219.0	1072.994	639.00
Residual	111	186.4	1.679	
Total	114	3405.4	29.872	

*** Estimates of regression coefficients ***

(Volume 100) = (Water filtered) - (Volume 10)

	estimate	s.e.	t(111)
Constant	0.833	0.191	4.36
Water filtered	3.390	0.193	17.59
Volume 0.9000	2.297	0.202	11.37
Volume 10	-2.2018	0.0788	-27.95
Volume 100	0	*	*

15.4 Sampling Exercise 7

15.4.1 Offset model

*** Summary of analysis ***

	d.f.	deviance	mean deviance	deviance ratio
Regression	1	2669.99	2669.990	1946.25
Residual	65	89.17	1.372	
Total	66	2759.16	41.805	

*** Estimates of regression coefficients ***

	estimate	s.e.	t(65)
Constant	3.9155	0.0400	97.94
Water filtered	-2.7892	0.0547	-51.03

15.4.2 Factor model

*** Summary of analysis ***

	d.f.	deviance	mean deviance	deviance ratio
Regression	2	570.42	285.209	223.68
Residual	64	81.61	1.275	
Total	66	652.02	9.879	

*** Estimates of regression coefficients ***

(Volume 10) = (Water filtered)

	estimate	s.e.	t(64)
Constant	1.863	0.105	17.79
Water filtered	1.566	0.111	14.14
Volume 0.9000	1.913	0.113	16.98
Volume 10	0	*	*