Health Protection Scotland

NHS National Services Scotland

A study of variation in serological markers for *Cryptosporidium* exposure over time associated with the introduction of enhanced filtration treatment of drinking water from Loch Katrine

Tender Reference: ENV/2005/5.6.8

Final Report – March 2011

Prepared by

Dr Kevin GJ Pollock and Dr Colin N Ramsay

Version 9_KP_CR280311

Table of Contents

Acknowledgements

The continued assistance and practical support of staff with the Scottish National Blood Transfusion Services (SNBTS) was invaluable in the successful completion of this study. In particular, the following staff at both Glasgow and Dundee donor centres were accommodating and diligent in recruiting donors into the study, without which it would have been impossible to perform the survey; in particular we acknowledge the assistance of Dr Moira Carter, Carol Garnett, Elaine Meikle and Trish Stevenson.

A number of project support staff within HPS were involved in the study and their input into the project is also gratefully acknowledged; Lesley Haldane, Jennifer Keatings and Elizabeth Burns.

Analysis of the serological samples was performed at the Scottish Parasite Diagnostic Laboratory (SPDL), in particular by Edwin Yip under the supervision of the late Professor Huw Smith, Director of SPDL.

Special thanks is recorded (posthumously) to Huw, who provided the specialist parasitology expert advice for the study and facilitated the serological testing at SPDL. Huw, together with Dr Catherine Benton (formerly of Scottish Water) and Dr Ramsay revised the original proposal for such a study. They, together with representatives of SNBTS and Matt Bower (Drinking Water Quality Regulator (Scotland) provided guidance and support to the project team throughout as the project steering group. Particular thanks is extended to Matt Bower for his patient assistance and perseverance.

The invaluable role of specialist statistical input to the design of the study and subsequent analysis is also acknowledged, particularly from Professor Chris Robertson and Dr Adam Wagner. We are indebted to Professor Robertson for his assistance in the interpretation of the findings.

Finally, the authors would like to acknowledge the Scottish Government for the funding of this project.

Abbreviations

ELISA	**Enzyme-linked-immunosorbent assay**
HPS	**Health Protection Scotland**
ID	**Identification**
Ig	**Immunoglobulin**
kDa	**kiloDalton**
PAGE	**Poly-Acrylamide Gel Electrophoresis**
PPR	**Percentage positive response**
SNBTS	**Scottish National Blood Transfusion Service**
Sp.	**Species**
SPDL	**Scottish Parasite Diagnostic Laboratory**
SW	**Scottish Water**
WTW	**Water Treatment Works**

Abstract

Introduction

Cryptosporidium is a protozoan parasite that causes the clinical illness cryptosporidiosis in humans, presenting as a diarrhoeal illness, which can be prolonged but is usually self-limiting. The infectious oocysts are carried and transmitted by a wide variety of animal species including humans. Zoonotic transmission occurs from animal carriers to man, as well as person-to-person transmission between humans. Infection can also be acquired indirectly through the exposure to contaminated food or water. Consumption of contaminated drinking water is a well-recognised risk factor for human cryptosporidiosis, presenting both as sporadic cases and as waterborne outbreaks. The first recorded waterborne outbreak in the UK occurred in Ayrshire, Scotland in 1988.

A number of additional high profile waterborne *Cryptosporidium* outbreaks occurred in Scotland and elsewhere in the UK, associated with unfiltered tap water consumption, failures of filtration treatment or post-treatment contamination of treated supplies. This lead to a Government enquiry, chaired by Sir John Badenoch. The report of this enquiry and subsequent (Bouchier) groups recommended a series of measures to reduce the risks of waterborne cryptosporidiosis. This, in turn, lead to significant investment in improved standards of water treatment, especially for unfiltered supplies. Scotland had several large examples supplying the Glasgow area (Loch Lomond, Loch Katrine).

Introduction of filtration to the treatment system for Loch Lomond was successful in significantly reducing the oocyst levels in the final water. This reduction in oocyst transmission via Loch Lomond supply drinking water was also associated with a reduction in the subsequent incidence of confirmed cryptosporidiosis in the Loch Lomond supply area, strongly indicating that a proportion of clinical cases had indeed been attributable to waterborne oocyst exposure in the pre-filtration era. Although this effect was considered

overall as a beneficial public health impact associated with introducing filtration, it also suggested a possible disadvantage in the longer term. Removal of chronic low level exposure to oocysts, by filtration of drinking water, might also paradoxically remove a low level stimulus to the production of protective anti-oocyst antibodies. This, in turn, might mean that consumers in such a supply area might become relatively more susceptible to *Cryptosporidium* infection in future, via other sources or following break-through of waterborne oocysts. A study was therefore designed to test this hypothesis.

Methods

The study was designed to detect evidence of an association between the introduction of filtration to the drinking water treatment system for Loch Katrine and changes in the level of antibodies to *Cryptosporidium* specific proteins (antigens) among residents of the relevant supply area. Blood donors living in the Loch Katrine supply area were recruited to provide blood samples for analysis (the Glasgow cohort). Donors were also recruited from another area having been supplied with filtered water consistently for many years, to act as a 'control' group for the Loch Katrine subjects. The population supplied by the Clatto reservoir in Dundee was selected (the Dundee cohort) as the 'control'. Serum from blood samples was tested for antibody to oocyst antigens on repeat occasions over a three-year period from 2006 to 2009 in both cohorts matched in time to control for the effect of any variation in exposure associated with seasonal variation. The period of study spanned before and after the new Loch Katrine filtration system became operational at Milngavie Water Treatment Works (WTW) in September 2007. The study was therefore designed to test for evidence of changes in the Glasgow donor antibody levels before and after filtration was introduced to Milngavie WTW. Furthermore, we compared the results to to those levels of antibody from the Dundee control population to check for any underlying changes or trends in exposure to *Cryptosporidium* associated with other sources of environmental contamination.

The study was ecological in type meaning that the results would apply at a population level but not at an individual donor level. Due to the complex before/after cross-over design of the study, the analysis of the results was complicated. Sophisticated statistical tests were therefore required to analyse the data to allow for the interactions between multiple possible exposure risk-factors over time using both univariate and multivariate analysis including liner and logistic regression.

Results

The main findings of the study are summarised:

- The average level of antibody detected in the blood donor serum samples in both Dundee and Glasgow cohorts increased over the period of the study. This was associated with the gradual ageing of the participants suggesting that antibody levels increased, consistent with an increased risk in exposure to *Cryptosporidium* oocysts over time.
- The trend for increasing average antibody levels for the Dundee cohort was consistent over time. However, the trend was reversed and the average antibody level decreased for the Glasgow cohort between the second and third sample periods. This coincided with the introduction of filtration at Milngavie WTW in September 2007.
- Statistical tests were used to investigate the multiple interactions which might explain the finding of the 'step change' in the antibody levels in the Glasgow cohort. These included variations in the donor demographics and frequency of other *Cryptosporidium* exposure risk variables. This analyses confirmed that the most important factor associated with the step change reduction in the Glasgow donor oocyst antibody levels was the introduction of filtration to the Loch Katrine supply in the relevant period.

- The reduction in average antibody levels in the Glasgow cohort proved to be only temporary. Between periods 3 and 4 of the sampling programme the underlying trend of increasing antibody levels over time among the Glasgow participants resumed. The average antibody levels in the Glasgow participants was effectively 'reset' from a level that was formerly higher than the Dundee cohort, to a level that was now lower. This reduction was attributable to the loss of chronic low level exposure to oocysts associated with the previously unfiltered Loch Katrine supply. Removal of this constant environmental 'top-up' stimulus revealed the underlying level of antibody response attributable to oocyst (or antigen) exposure from other sources, which appeared to be lower in the Glasgow participants compared with the Dundee cohort.
- The findings relate to a self-selected sample of the population (blood donors) who represent a relatively fit and healthy group of adults. The study design did not permit assessment of the whole population exposed to Loch Katrine sourced drinking water due to the practical constraints of obtaining representative blood samples for all age groups. The results of this study may therefore not be generalisable to the whole population equally. There may be a differential impact from removing low level environmental exposure to *Cryptosporidium* on those individuals in the population who may already be more vulnerable to infection (e.g. young children and the immuno-compromised).

Conclusions

The study hypothesised that there would be a reduction in the anti-*Cryptosporidium* antibody levels in the population supplied with drinking water from Loch Katrine associated with the introduction of filtration at Milngavie WTW in September 2007. The 'null hypothesis' considered that there would be no effect. The association of filtration and a reduced average oocyst antibody level in the relevant population was confirmed. The results allowed rejection of the null hypothesis.

This reduction in antibody levels appears to be only temporary and the underlying trend of a steady increase in the average levels of such antibodies over time was strong enough to overcome the initial reduction associated with the elimination of waterborne oocysts from the Loch Katrine sourced drinking water. This indicates that people continue to encounter low level exposure to *Cryptosporidium* oocysts or their proteins from a variety of sources and that this continued exposure provides a low level stimulus for the immune system, which may confer protection against future large dose exposure to oocysts and mitigate against developing cryptosporidiosis.

Although there was a temporary reduction in the levels of oocyst antibody among the Glasgow cohort, this did not persist. Hence, it is unlikely that there will be a sustained long-term increase in the vulnerability of the adult population in the Loch Katrine supply area to *Cryptosporidium* infection. The loss of the low level exposure from unfiltered drinking water could mean however that young children might have delayed exposure to oocysts and their proteins and might therefore have relatively delayed development of anti-*Cryptosporidium* antibodies compared to children who consumed unfiltered Loch Katrine drinking water in the past. This deficit is likely however to be compensated for over time by exposure to oocysts and/or proteins from other environmental, human or animal sources.

Recommendations

The clinical significance of the temporary reduction in oocyst antibody levels in the Glasgow population associated with introducing filtration to Loch Katrine supplies should be investigated to assess if there is any detectable impact on the incidence of clinical infection and if present, whether this is also temporary. The incidence of clinical cryptosporidiosis should therefore be assessed before and after September 2007 to detect evidence of any changes associated with water consumption or other recognised exposure risk factors.

Continued vigilance will be required to ensure that the newly introduced filtration system continues to perform optimally. Should the system fail to remove *Cryptosporidium* oocysts for any period, some consumers will remain at risk of infection (as before) from such exposure. Monitoring of oocysts in both raw and final water should therefore continue in order to identify any periodic increase in the level of risk attributable to waterborne oocysts.

1 Introduction

Health Protection Scotland (HPS) receives reports of between 500 and 700 laboratory-confirmed cases of cryptosporidiosis each year. Characteristic symptoms of cryptosporidiosis are profuse, watery diarrhoea, often accompanied by bloating, abdominal pain and nausea or vomiting. While illness can last for 2–3 weeks, it is normally self-limiting, although a recent study has shown an association with significant long-term health sequelae such as reactive arthritis [1]. Moreover, people with severely compromised immune systems, particularly with reduced T-cell counts may develop severe chronic diarrhoea or atypical gastrointestinal infection in the absence of immunotherapy, which may prove fatal [2].

Cryptosporidiosis is caused by one or more species/genotypes of the genus *Cryptosporidium*, a protozoan parasite which infects a wide variety of animal species including humans. Asymptomatic carriage is common. The most important human pathogens, which account for the majority of human illness in the UK, are *Cryptosporidium hominis* and *Cryptosporidium parvum*.

Drinking water contaminated with *Cryptosporidium* oocysts is an internationally recognised risk factor for human illness [3-5]. *Cryptosporidium* sp. contamination of drinking water can arise from a variety of sources [6] including infected humans, livestock and feral animals present in the catchment via the contamination either of raw water or as post-treatment contamination with oocysts. *Cryptosporidium* oocysts can remain infectious in the environment and water for prolonged periods (several months) and are resistant to most disinfectants used to treat drinking water. Inadequate treatment of drinking water can permit infectious oocysts to be transmitted to susceptible consumers of that water, posing a risk to public health [7-10].

Outbreaks of cryptosporidiosis associated with the contamination of drinking water supplies are well-described in Scotland, elsewhere in the UK and internationally [3, 7, 8].

The first recorded waterborne outbreak in the UK occurred in Scotland, in Ayrshire in 1988 [7].

Current data on *Cryptosporidium* infection in Scotland are inadequate in predicting the human health impact of introducing improved standards of water treatment to reduce waterborne *Cryptosporidium*, by removal of oocysts using efficient filtration-based technology. Retrospective analysis of *Cryptosporidium* infection rates in populations receiving drinking water, treated to an improved standard of oocyst removal, has demonstrated some impact with evidence of reduced rates of infection [5, 10]. However, such retrospective studies depend on the quality of routine public health case investigations carried out at local level, which may vary in the quality of information on potential risk factors for exposure to *Cryptosporidium*.

Variations in the significance of other prominent risk factors (e.g. foreign travel, direct contact with animals) may also obscure a low-level effect associated with a change in oocyst exposure risk via local drinking tap water. Ideally, an assessment of change to oocyst exposure would be determined on the basis of numbers exposed in the entire population at risk, rather than via laboratory-confirmed cases of cryptosporidiosis, who can only ever represent a self-selected subset of the whole population, exposed to drinking treated tap water. Testing of human blood samples (serum) for evidence of exposure to oocysts, in the form of specific antibodies (immunogloblulins) to oocyst proteins, is one method of assessing how common exposure is in a population. It also offers a method of tracking changes in oocyst exposure over time associated with changes in exposure to factors such as drinking tap water.

However, the relationships between *Cryptosporidium* exposure, infection and specific antibody formation (seropositivity as a measure of *Cryptosporidium* exposure) and the actual levels of *Cryptosporidium* oocysts in any drinking water supply are complex. Low levels of *Cryptosporidium* oocysts have been detected in 65-97% of surface-water supplies throughout the world, suggesting that most populations may be at risk for

waterborne infection [11]. A US study reported that consumption of 'surface-derived drinking water, treated with conventional filtration and chlorination and meeting all water quality standards, may carry an increased risk of waterborne *Cryptosporidium* infection' compared with groundwater [13]. Furthermore, several serological studies conducted in pairs of cities with different types of water supplies found elevated *Cryptosporidium* antibody levels (serological responses) in users of surface-water versus groundwater supplies. This suggests that the risk of exposure to oocysts from surface-derived drinking water may be higher than for groundwater [12-14]. Paradoxically, recent studies also suggest that relatively strong serological responses (>20% of the positive control response), possibly resulting from chronic low-dose waterborne and/or foodborne oocyst exposure, may confer some protective immunity to clinical infection with *Cryptosporidium* [17, 18].

Responses by the Water Industry (and public health agencies) to *Cryptosporidium* contamination of drinking water have focussed in Scotland on establishing effective multiple barrier water treatment systems, using a variety of methods (slow sand filtration, rapid gravity filtration with coagulation and membrane filtration) in an effort to eliminate this waterborne pathogen from drinking water supplies. This has resulted in a situation where the majority of drinking water supplies, especially in major urban areas in Scotland, have effective forms of water treatment capable of significantly reducing the *Cryptosporidium* load in final drinking water. One of the few exceptions to this was the Loch Katrine distribution system, which supplies the Glasgow and Clyde population (approximately 700,000) with drinking water. Prior to September 2007, treatment of the Loch Katrine raw water supply consisted of only micro-straining and chlorination. After September 2007, all drinking water supplied from Loch Katrine underwent enhanced filtration, following a major investment in new water treatment facilities by Scottish Water.

Given evidence of a possible relationship between continuous low level oocyst exposure and titres of protective antibody in humans, it was postulated that the continuous low level oocyst exposure of Loch Katrine water consumers (prior to enhanced water

treatment) might be associated with a higher level of population immunity to infection with *Cryptosporidium* than other populations [13-15]. This provided an opportunity to conduct a study on the impact of introducing the enhanced standard of water filtration on the background level of detectable antibodies to *Cryptosporidium* in consumers of Loch Katrine drinking water.

If continuous low-level exposure to *Cryptosporidium* oocysts (or their antigens) via contaminated drinking water results in the maintenance of partial protective immunity against infection in such exposed humans, then an unintended consequence of introducing filtration to the Loch Katrine supply might be a reduction in the background level of 'herd immunity' to *Cryptosporidium.* A reduction in such background immunity might then render people more susceptible to infection from (any) subsequent oocyst exposure. Hence, a future failure of enhanced water treatment might paradoxically result in an increased risk of outbreaks of clinical cryptosporidiosis amongst populations who by then normally receive drinking water, which is free of *Cryptosporidium* [15].

To this end, we therefore proposed a study to investigate the relationship between levels of *Cryptosporidium* antibody levels (seropositivity) and exposure to drinking water, before and after the introduction of enhanced water treatment technology to Loch Katrine water supplies at Milngavie Water Treatment Works (WTW), in 2007. The principal aim of the study was to try and detect any changes in background seropositivity among Loch Katrine water consumers associated with the change in water treatment.

One method for investigating the incidence of disease is to measure the extent of exposure to the causal organism by testing for a specific antibody (serological) response to that pathogen. There have been few serological studies of *Cryptosporidium* exposure in human populations but antibody responses to the parasite antigens (in general) have been shown to be consistent and of sufficient intensity to represent dependable measures of exposure to the organism [17, 18]. Serological studies of oocyst exposure have focused on the immunogloblulin G (IgG) responses to the 15/17-kDa and 27 k-Da protein groups.

Responses to these two parasite ‘antigens’ appear to be specific to *Cryptosporidium* infection and therefore these provide a robust marker of exposure to oocysts. Clinical infection (where someone develops clinical symptoms of illness) usually elicits an antibody response to these two antigen groups, that peaks 4-6 weeks after infection [14]. The 15/17-kDa marker declines to baseline levels 4-6 months after an infection, while the 27-kDa marker remains elevated for 6-12 months [15]. The latter antigen therefore provides an indicator of past exposure over a longer time period.

We therefore investigated the seroprevalence (the proportion of the population with evidence of specific antibodies to *Cryptosporidium*) of the 27-kDa antigen group in people living in two areas of Scotland with different sources of drinking water and different standards of water treatment; Loch Katrine in Glasgow (un-filtered) and Clatto reservoir (conventionally filtered water) in Dundee (for comparison). The study commenced in 2006. In September 2007, coagulation and rapid gravity filtration was introduced at the Loch Katrine WTW. The study continued to assess the seroprevalence of *Cryptosporidium* antibodies among the study participants until October 2009. The primary objective of the study was to determine whether there was evidence of an association between the seroprevalence of the 27-kDa marker and the type of treatment (filtered vs unfiltered) used on drinking water water supplies.

The study plan was to recruit a cohort of blood donors in both Dundee and Glasgow and to follow these individuals over the duration of the study. The intention was to provide a longitudinal serological profile for each donor, allowing determination of antibody kinetics over time. This method would also allow estimation of the changes in the proportion of the respective cohorts who demonstrated evidence of exposure over the study period, related to their increasing age as individuals. The study method therefore enabled a multi-faceted analysis of evidence for *Cryptosporidium* exposure, an unparalleled opportunity to study the seroprevalence of *Cryptosporidium* in two Scottish study cohorts.

1.2 Aims and objectives

Aim

The primary aim was to determine evidence of association between the seroprevalence (the level of antibody response) to the 27-kDa *Cryptosporidium* antigen and exposure to unfiltered or filtered drinking water in a sample of consumers.

Objectives

The objectives of this research study were to:

1. Determine the demographic profile of the Glasgow and Dundee donor cohorts, identify their *Cryptosporidium* exposure risk profiles and compare the cohorts to identify any significant differences, which might influence the study findings.
2. Establish the background levels of seropositivity to *Cryptosporidium* in the two donor cohorts (Glasgow and Dundee) over the study period 2006 to 2009.
3. Determine if there is a significant difference in seropositivity associated with differences in drinking water treatment (filtration) between Dundee and Glasgow donors, before and after September 2007.
4. Determine if there is any appreciable difference in trends over time in the levels of seropositivity between Dundee and Glasgow donors associated with the introduction of filtration to the Loch Katrine supply.
5. Ascertain whether fluctuations in Loch Katrine water and Clatto water are relevant in relation to variations in evidence of seropositivity ot oocyst exposure.

2 Methodology

2.1 The selection of study sites and populations

The study design is an 'ecological' type of epidemiological study. The main characteristic of ecological studies is that evidence of association between exposure and a health effect is assessed daily at a population level (not the individual person levels). Exposure is assessed for the population collectively, via measurement of the proportion of the population who demonstrate evidence of *Cryptosporidium*-specific antibodies and who live in either the area supplied by Loch Katrine, or the Clatto reservoir. Exposure and effect are therefore assessed only at population level, not on the basis of quantifying exact exposure of individuals and correlating this with their personal serological results. The consequence of this method is that if an association is demonstrated, it applies only at population level and cannot necessarily be extrapolated to the level of risk at an individual consumer level. Nonetheless, ecological studies are well-recognised as appropriate for the investigation of population exposures to common factors, such as drinking water [10].

This particular study has a number of refinements. The study uses the comparison between levels of seropositivity to *Cryptosporidium* among consumers of filtered water (Clatto) with consumers of unfiltered water (Loch Katrine) to investigate evidence of any association between seropositivity and water filtration, albeit confounded with location and population. Moreover, the study compares consumers of Loch Katrine water before and after filtration was introduced to the same water source i.e. unfiltered vs filtered water consumption in the Loch Katrine cohort. This provides a complex comparison matrix designed to maximise the ability to detect an association, if one exists. Such complex refinements to ecological studies are unusual in work on *Cryptosporidium* and this study is the only one to date involving *Cryptosporidium*.

The study was conducted in two distinct and geographically separate locations in Scotland, centred around Glasgow and Dundee. Both areas are served by surface-derived

drinking water sources. Clatto reservoir, the supply for Dundee, has had coagulation, rapid gravity filtration and disinfection for many years, long before this study commenced. In contrast, the treatment of Loch Katrine reservoir water used to supply Glasgow (Milngavie WTW), was only upgraded in September 2007 to a specification matching that of Clatto.

The study received approval by the Multi-Research Ethics Committee for Scotland. Investigating the seroprevalence of antibodies in a population requires access to blood samples. The scope of the study precluded the collection of blood samples specifically for the study itself (the cost and complexity of such a study would have made it impractical). It is also not possible to sample an entire population, hence a study will conventionally try to use a representative sample as a proxy for the whole population. Hence, blood donors were identified as the sample group and agreement to use samples of their donated blood was obtained from the Scottish National Blood Transfusion Service (SNBTS). In view of the use of human blood samples, approval was sought and received from the Multi-Research Ethics Committee for Scotland.

Blood donors from clinics in Glasgow and Dundee were recruited to form the two respective cohorts. All subjects were informed of the aims of the study and asked to provide consent to taking part in the study for the duration of the project. Each donor was asked to complete a questionnaire requesting personal information and their history of exposure to potential risk factors for *Cryptosporidium*. Prior to the relevant blood donation session, blood samples were obtained from each donor and sent in heparinised sample tubes to the Scottish Parasite Diagnostic Laboratory (SPDL) for storage and analysis.

An advantage of blood donors is that they are screened for recent infectious diseases and so factors which might interfere with their immune response and hence potentially affect their immunological profile, are adequately minimised. However, blood donors are also a self-selected group of relatively fit and healthy adults, ranging in age from 18 to 60 and therefore represent a sub-set of the general population. The study populations in Dundee and Glasgow were both drawn from blood donors to ensure

comparability, hence the findings of the study must be interpreted in light of the sample selected.

The statistical power of the study to detect a real difference in seropositivity associated with filtered verus unfiltered water was related to the background levels of antibodies to *Cryptosporidium* in the respective populations. The higher the background level of antibody in the Clatto water consumers, then the higher the antibody levels would need to be in the Loch Katrine consumers for a statistically significant difference to be detected between consumption of filtered vs unfiltered water. In Dundee, 300 blood donors were to be recruited and in Glasgow, 700 blood donors. The reason for the disparity was to ensure that there was more information in the population where the filtration system was to be introduced.

It was calculated that if 5% of the Dundee donors had a positive serological response then the study would be able to identify significant differences between Dundee and Glasgow donors, with 95% power providing at least 11.8% or more of the Glasgow donors exhibited a positive serological response. If 30% of the Dundee donors displayed a positive serological response, then the study would identify significant differences between Dundee donors and Glasgow donors, with 95% power, if 41.9% or more of the Glasgow donors also produced a positive serological response.

The power to detect changes in antibody levels pre- and post-filtration in Glasgow was calculated on the assumption that McNemar's test would be used. However, we also had to allow for the fact that we might lose some of the original donors (attrition, repatriation etc.) and therefore obligate recruitment of replacements. Assuming 30% of the donors were positive at baseline (first blood test) then a sample of 700 people has a power of 95% probability in detecting a difference of at least 4.5% in the underlying seropositivity rates between the populations (i.e. from 30% to 25.5%) if 10% of the donors change. If 20% of donors change, then the detectable difference would be 5.8% (i.e. from 30% to 24.2%) and if 30% of donors change the detectable difference is 7.5%, i.e. from 30% to

22.5%. In other words, the more that original donors are lost from the study, the bigger the difference there would have to be between seropositivity associated with drinking filtered versus unfiltered water in the two populations for a statistically significant difference to be detected, if it existed. In summary, a sample of 700 donors with repeat measurements on all 700 will give high power (95%) to detect differences of at least 10% in seropositivity between the two populations sampled.

It was planned that these 1000 blood donors would give a 'baseline' sample and 3 further follow up samples, two of which would be after the new filtration system was introduced at Loch Katrine. Thus there would be 2 measurements pre-filtration and 2 post-filtration for each donor in both Dundee and Glasgow to allow for changes in other exposures and changes over time.

2.2 Western blotting

Serological responses appear to develop after exposure to *Cryptosporidium* oocysts, even in the absence of clinical illness [16]. Such antibodies produced in response to *Cryptosporidium* exposure are readily detectable. The use of serological assays as a marker for exposure reduces the potential under-ascertainment associated with reliance on only a reported history of *Cryptosporidium* infection. The method is based on the analysis of blood samples, to identify the presence of antibodies specific to *Cryptosporidium* (seropositivity). The antibody testing involves Western Blot analysis and was performed at SPDL under the supervision of the late Professor Huw Smith.

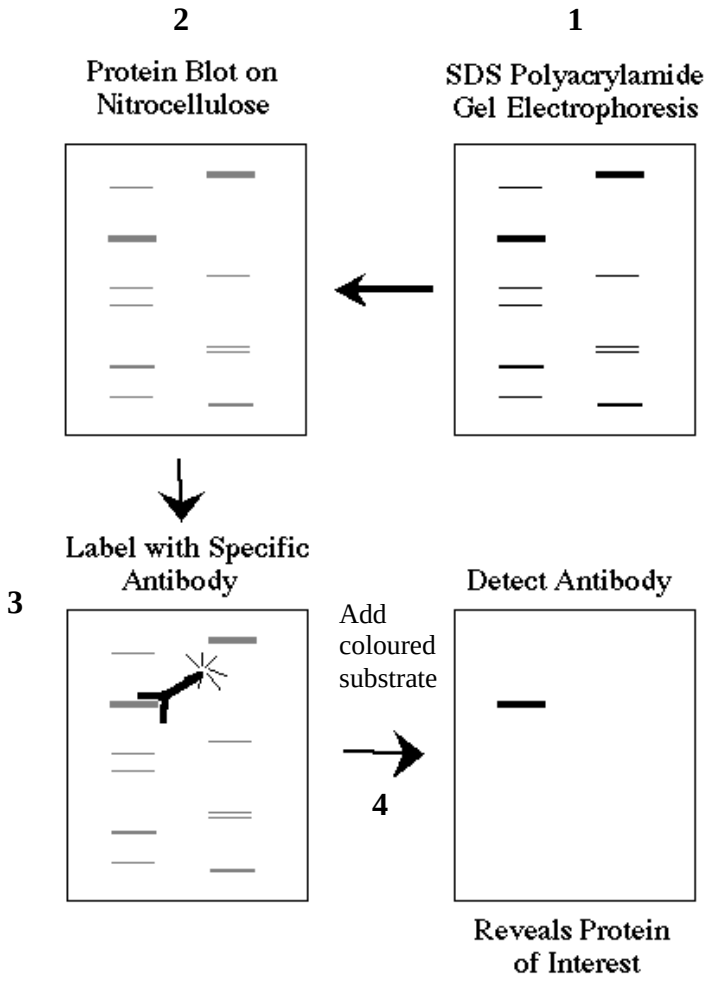
The Western blot method allows investigators to determine the molecular mass of an antigen and to measure relative amounts of that antigen present in different serum samples. The process is summarised:

1. Proteins, in this case from the *Cryptosporidium* parasite, are separated by poly-acrylamide gel electrophoresis (PAGE). Essentially, this is a molecular sieve. As

the proteins are negatively charged, a negative electric current repels the proteins and the lowest molecular weight proteins migrate through the gel quickest while the largest proteins remain closer to the origin of application.

2. The proteins are transferred to a sheet of special blotting paper (blot) called nitrocellulose, though other types of paper, or membranes, can be used. The proteins retain the same pattern of separation as they had on the original gel (Figure 1).
3. The blot is incubated with a generic protein (such as milk proteins), which binds to any remaining vacant binding sites on the nitrocellulose. Material containing an antibody is then added to the suspension which is able to bind to its corresponding antigen on the blot. In this study, human serum samples from blood donors were added as the source of antibody. Each sera sample may or may not contain *Cryptosporidium*-specific antibodies (indicating past exposure to oocysts), which bind specifically to the protein of interest – the 27 kDa protein. A second antibody is then added to this membrane, which specifically binds the *Cryptosporidium*-specific antibody, if it is present. The secondary antibody has an enzyme (e.g. alkaline phosphatase or horseradish peroxidase) attached to it.
4. The location of the antibody on the blot is then revealed by incubating it with a colourless substrate that the attached enzyme converts to a coloured product deposited at the site of secondary antibody binding. This colour product can then be visualised and photographed. The band densities are then optically measured and compared with a positive control. This provides the measure of how strong an antibody response that particular blood donor has to the *Cryptosporidium* antigen.

Figure 1. Schematic of Western blot methodology.



2.3 Interpretation of Western blot results

The Western blot method compares each blood donor serum sample against a positive control. Ideally, the positive control would be derived from sera which originated from cases of human clinical illness (i.e. someone who had been infected with *Cryptosporidium*). Unfortunately, this study did not have access to sufficient human case-derived sera material. Hence, the study relied on rabbit-derived positive controls (serum from a rabbit which has been immunised with a *Cryptosporidium* lysate of the parasite which generates/produces a standardised response). This form of calibration is recognised

as an acceptable method and has been used in comparable studies. The positive control response is taken to represent a value of 100% seropositivity. Blood donors may range from having non-detectable 27-kDa antibody to *Cryptosporidium* (0% positivity) to a detectable result which equates to the positive control response (100% positivity) or may even exceed that of the positive control (e.g. 310% in one instance).

The use of Western blot technology for analyses of serological (antibody) responses to *Cryptosporidium* has previously been used by other investigators [15, 17, 18]. These studies have used the 15/17- and 27-kDa protein groups and have provided estimates of the prevalence of prior infection and identified risk factors for parasite transmission [12-14, 17, 18]. Serum samples in this study were analysed by Western blot, to measure the IgG serological response to the 27-kDa *Cryptosporidium* protein only. The 15/17-kDa complex was not tested for a number of reasons. The 15/17-kDa marker declines to baseline levels observed prior to the infection 4-6 months after infection. As the study collection periods were spaced months to years apart and the 27-kDa antibody response is the longer lasting one, we did not see any benefit in collecting data on the 15/17-kDa complex. Furthermore, the logistical complexities of this study would have been greatly added to, if another level of analysis was needed for data on the 15/17-kDa complex.

The Western blot assay was validated before testing of participant samples. In order to calibrate a negative serological response (no antibodies), a pool of 10 human sera from the SPDL negative serum bank was used; these were known to be serologically negative to various human parasites including *Cryptosporidium*. In order to calibrate a positive serological response, a pool of 10 human sera from the SPDL *Cryptosporidium* positive serum bank was used; these were collected from known human cases excreting *Cryptosporidium* oocysts and were therefore sero-positive to *Cryptosporidium* oocyst antigens. There is only a very small pool of positive sera samples held locally for *Cryptosporidium* cases, hence these had to be used sparingly. Similar validation was performed on the rabbit control sample i.e. for negative control serum, serum from a

specific pathogen free rabbit, which was sero-negative to *Cryptosporidium* oocyst antigens, was used. Conversely, serum from a specific pathogen free rabbit, which was immunised with soluble *Cryptosporidium* oocyst antigens in adjuvant, was used.

The alternative ELISA technology for antibody detection was explored as a pilot project but not used for the main study due to financial and other constraints (Appendix 1).

2.4 Statistical methods

The study generated a large and complex dataset. Analysis of this was itself both complex and reliant on sophisticated techniques and advanced statistical input. These ranged from univariate techniques, where the distribution of each of the epidemiological exposure questionnaire variables and serological measurements are compared in relation to the levels of one explanatory variable, to multivariate methods, where a number of explanatory variables are simultaneously considered. The comparison of the distribution of serological responses between the Glasgow and Dundee blood donors is an example of a univariate analysis. If serological response is influenced by age and gender, then if the age and gender distributions themselves are different between respondents in Glasgow and Dundee, then the univariate comparison may be biased. This bias can be corrected by performing multivariate analysis.

The multivariate technique used to take into account the effect of a number of explanatory variables on one response variable is ‘regression analysis’ and in this report, linear regression, logistic regression and poisson regression are all used. The choice of the type of regression model is dictated by the distribution of the response; for continuous response variables, linear regression is used (e.g. serological response); for binary responses (e.g. yes/no responses) then logistic regression is used, and where the response is the count of the number of events in a fixed area or volume, poisson regression is used (e.g. value of serological results in the study period).

Essentially, the data-set was analysed in three steps:

1. analysis of donor epidemiological questionnaire responses (see Appendix 2)
2. analysis of serological responses and
3. analysis of both elements combined to determine the most important factors associated with the serological response.

Analysis of the donor questionnaire responses provided 'risk profiles' of donors and allowed comparison of how these differed between donors from the two geographic cohorts over the time period studied. Descriptive analysis was used to define the characteristics of blood donors and the distribution of donors by water source, before and after the introduction of enhanced physical treatment at Loch Katrine WTW. The statistical significance of associations between 'categorical' variables (e.g. age) and city were investigated using **chi-squared** or **Fisher's exact tests**. Fisher's exact test is an alternative test to chi-squared, used where the lowest 'expected' values in a 2 x 2 table are very small (< 2).

Means and standard deviations were used to provide summary statistics for donor serological responses by city and collection period. Such techniques can only be used to 'infer' differences, if there are any. In order to confirm a true difference, regression modelling was used; a technique called '**mixed effects modelling'**. The serological responses are considered using two different methods. In the first method, the serological responses are calculated as a **binary** valued variable; the serological response is considered 'positive' if it is above a range of pre-agreed thresholds (e.g. a threshold of 0, 10, 20, 30, 40 or 50% positivity, compared with the 100% positive control) and 'negative' if otherwise. Hence, setting the threshold at 0% means any sample with any antibody response is considered 'positive'; setting it at 100% means only samples that matched the positive control sample response would be deemed 'positive'. The range of 'positive' thresholds was chosen to allow a sensitivity analysis of the results. Because a rabbit serum model is used, the actual positive control value (deemed 100%) is in effect arbitrary and not of itself

a definitive measure of antibody positivity. The assay uses comparisons and relative responses hence the need for the sensitivity analysis.

In the second statistical method, the serological responses are modelled directly using a **linear** mixed effects model. For both statistical models, a statistical construct (a 'contrast') is used to test for differences in serological response between collection periods, water sources (cities as a proxy) and differences that are due to a combination of these two factors. Where 'Dundee' and 'Glasgow' are used as proxy terms, these denote the donor cohorts supplied with drinking water from Clatto reservoir and Loch Katrine respectively.

Analysis of serological responses and risk profiles in combination, was investigated by using **linear mixed effects model** analysis to quantify the characteristics of Glasgow donors compared with the Dundee cohort.

The reason that **mixed effect regression models** are used is because there are repeated observations on the majority of the study participants over time. As designed, each participant should have had serological measurements at four time periods. These four measurements cannot be considered to be totally independent of each other, and are expected to be correlated with each other since they are taken from the same person. We would expect that measurements taken at the same time period but on different people would be independent. To take this 'within-person correlation' into account, a mixed effect model is required. Although the study was based on a power calculation using McNemar's test, this was not used in the analysis, as more complex statistical methods were required. These methods make full use of the data at each period and were also needed to take into account the inclusion of new donors (recruited to replace those lost to attrition) after the baseline period.

Data reflecting the actual levels of oocyst contamination in the raw water supplies was also collected to verify that the risk of *Cryptosporidium* exposure did exist after all. Data on turbidity (a measure of soil contamination in water) and oocyst counts, obtained from Scottish Water, were also analysed for Clatto and Loch Katrine. To investigate the

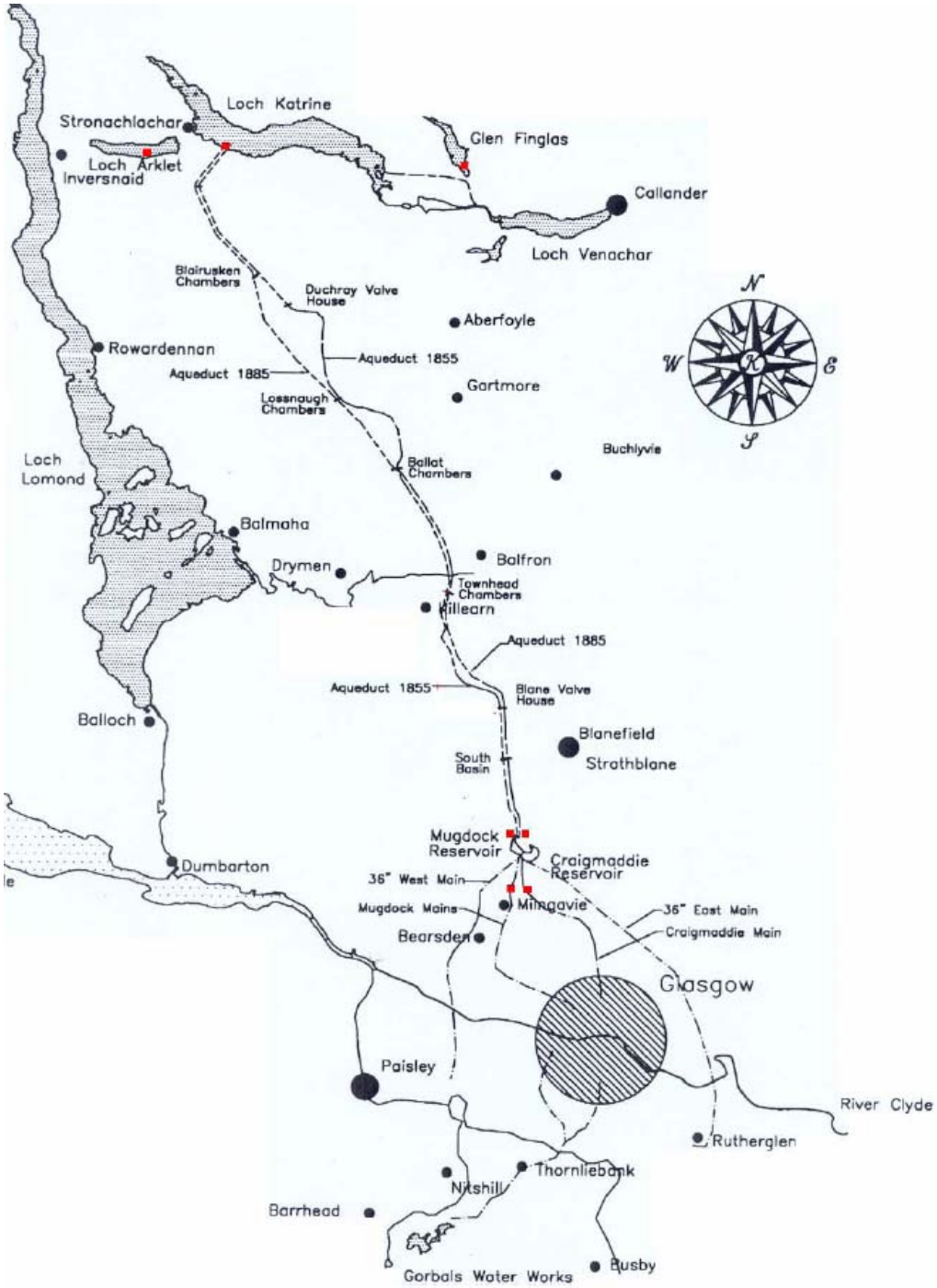
effect of filtration on these water quality measurements, monthly averages were aggregated and a **generalised additive Poisson model** was used for further analysis. The statistical models used in the study were selected because they specifically support the use of multi-dimensional data, as in this dataset.

Unless otherwise stated, we used R, version 2.80 (http://www.R-project.org) for the statistical analysis, while Minitab statistical software, version 14 (www.minitab.com) was used for the collation of information. In general, a significance level of 5% (i.e. $p < 0.05$) has been used for all analyses and should be assumed unless we state otherwise.

2.5 Oocyst counts in 'final' drinking water

Rather than relying on an assumption about the risk of exposure to *Cryptosporidium* oocysts via drinking water (as is the case in a number of other comparable studies of serology profiles [14,18]), in this study we also analysed data on oocyst counts detected in the relevant water supplies. Water supplies are regularly monitored for the presence of *Cryptosporidium* oocysts. In order to compare the levels of oocysts before and after introduction of filtration at Milngavie WTW, a schematic for the Loch Katrine water supply is depicted (Figures 2, 3 and 4).

Figure 2. Layout of Loch Katrine supply to Glasgow area.



The Dundee supply receives water via Clatto reservoir so there is only one set each of measurements for raw (untreated) and final (treated) water. For the Loch Katrine supply however, prior to September 2007, there were three different sites where sampling occurred for raw (untreated) water; Royal Cottage; Craigmaddie; Mugdock and two sites for final (treated) water – Mugdock WTW and Lambhill (figure 3). These were averaged to provide single mean concentrations for raw and final water. After the introduction of filtration at the Loch Katrine WTW, the sampling points changed with only one raw water (untreated) sampled at the inlet to the new WTW (this was a different site from the three raw water sampling sites prior to September 2007) and one final (treated) water at the outlet from the new WTW (figure 4). This introduced a complication to the comparison of before/after raw water quality from the Loch Katrine supply. This is discussed further in the Results section.

Figure 3. Schematic of water distribution from Loch Katrine to Glasgow consumers before filtration was installed (before September 2007).

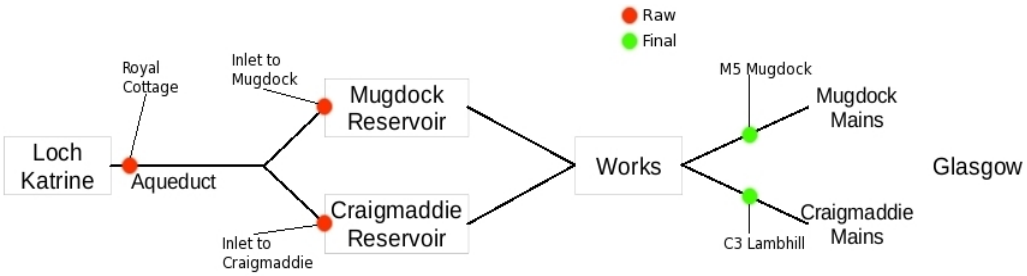
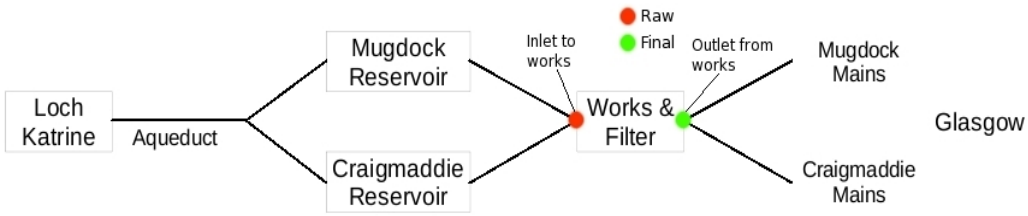


Figure 4. Schematic of water distribution from Loch Katrine to Glasgow consumers after filtration was installed (after September 2007).



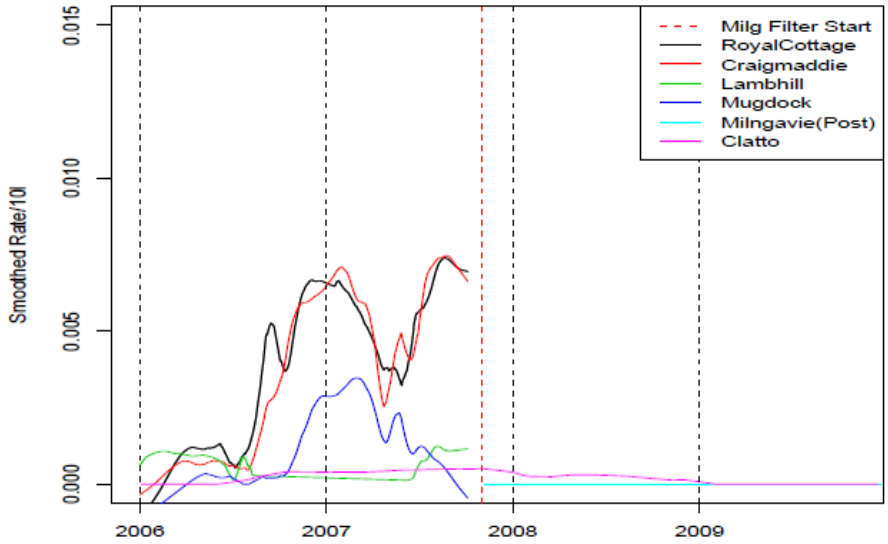
3 Results

3.1 Raw and final water oocyst counts

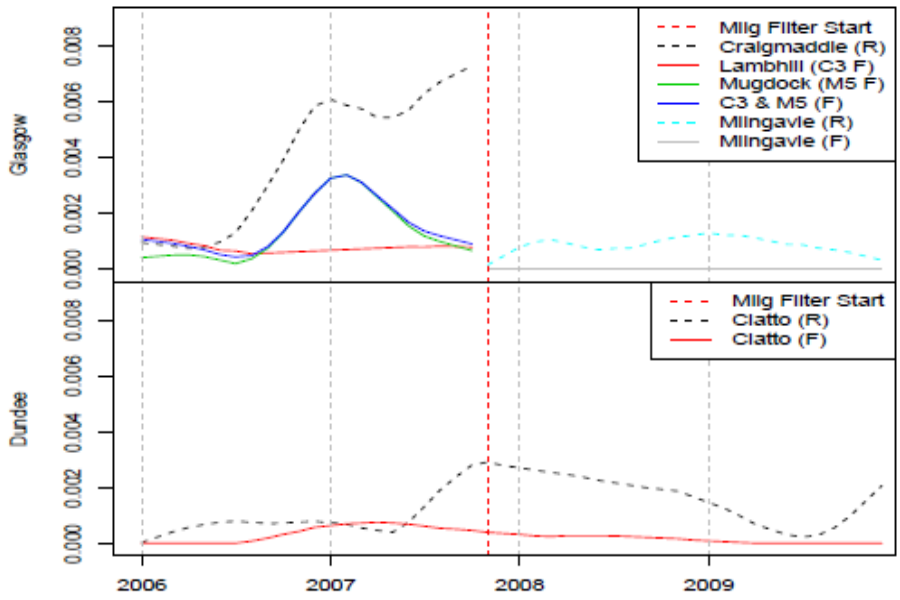
In order to analyse oocyst counts statistically before and after introduction of filtration at Milngavie WTW, a Generalised Additive Model (GAM) was used. GAMs allow the data to be smoothed out, which in turn helps to clarify whether there are trends associated with the data. Such trends must be modelled before appropriate comparisons can be made before and after introduction of filtration. Detection of oocysts in water is complex and dependent on an adequate (large) volume of water being tested. Large sample volumes are used for oocyst detection to reduce scope for sampling bias and therefore a Poisson model is used, which models the 'rate' (oocysts per litre) rather than the absolute number of oocysts observed. Monthly averages of the oocyst detection rates were computed and then smoothed: the resulting smoothed oocyst rates are shown in Figure 5.

Figure 5. Smoothed monthly average rates of oocysts per 10 litres. The vertical line indicates when the filtration system was installed in Glasgow. Raw and final water are denoted by 'R' and 'F' respectively (raw sites have dashed lines).

(a) Individual sampling sites



(b) Average of sampling sites (Craigmaddie (R) denotes Royal Cottage/Mugdock/Craigmaddie)



Unfortunately, the sampling locations for raw water for Loch Katrine sourced water changed before and after filtration implementation. This makes an appropriate analysis before and after filtration problematic. Generally, the rates of oocysts detected in final (treated) water were lower than those in raw (untreated) water for both cities, over most of the study period. This is consistent with the expected effect of treatment on raw water, which aims to reduce turbidity and improve microbiological quality of final drinking water.

Introduction of the enhanced filtration system at Milngavie WTW post-September 2007, reduced the level of oocysts in final water to unrecordable levels. However, this appears to have coincided with a decrease in oocyst counts in raw water. The move of the sampling point for raw water sampling at Milngavie WTW appears to have been associated with a marked change in the detected level of oocysts in raw water. In the pre-filtration era, sampling was at Royal Cottage, a location near the take-off point of water from Loch Katrine into the main supply aqueduct (marked as a square South of the Loch in figure 2). The second and third points were at the inlets to the Mugdock and Craigmaddie reservoirs. The Royal Cottage sample point was much closer to the sources of contamination of raw water, e.g. run-off into Loch Katrine from the surrounding catchment area, than the new raw water sample point at the new Milngavie WTW, in the post-filtration era. Hence, it is not surprising that oocyst counts (rates) were higher, in the earlier era given that the raw water sampling point (after filtration was introduced) was downstream from the two raw water holding reservoirs. These reservoirs are likely to have resulted in a degree of settlement of oocysts. There may also have been other factors which resulted in a true reduction in oocyst counts post-September 2007 e.g. changes to livestock rearing in the catchment.

Of note was the observation that the raw water oocyst concentration at Milngavie WTW in the post-September 2007 period was comparable to that of the Clatto reservoir in the pre-September 2007 period. The raw water sampling changes have therefore effectively

become a confounding factor for the filtration system, as raw water (post-filtration implementation) had a lower average rate of oocysts per 10 litres than previously. In fact, the oocyst count in raw water in Glasgow after September 2007 had lower oocyst rates than the final water from the old WTW, prior to the filtration system being implemented (Table 1); 0.00085/10 litres vs 0.00132/10 litres.

Table 1. Rate of oocysts in the raw and final water before and after the installation of filtration system in Milngavie.

Period	**Concentration of oocysts/10 litres**		
Dundee	**Raw**	**Final**	**% Reduction**
Pre-filtration	0.000850	0.000439	48
Post-filtration	0.001979	0.000115	94
Glasgow			
Pre-filtration	0.004219	0.001320	69
Post-filtration	0.000855	0.000000	100

The importance of the raw water oocyst concentrations does not lie in the magnitude of the absolute values but rather in providing proof that oocysts were present in the raw water prior to filtration. The critical measure of potential exposure to waterborne oocysts was the concentration in the final treated water at both Clatto and Milngavie WTW. The presence of oocysts (and/or their proteins) in the final treated water is what provides the immune system stimulus that the study was designed to detect.

The detection of oocysts in water destined for human consumption indicates that there is a risk of exposure to oocysts if contaminated water is consumed by susceptible (non-immune) humans. The higher the oocyst rate in final (treated) water, the more likely

that consumers will be exposed to oocysts via this route and it follows that they will have an increased probability of developing antibodies to oocyst proteins. This continuous, low-level environmental exposure to oocysts could have the effect of continuously stimulating antibody production to oocysts, which in turn may confer protection enabling the individual water consumer to resist symptomatic infection with *Cryptosporidium*. Consumers are only likely to be exposed to final water not raw, untreated water (unless the WTW was completely by-passed). Variations in the waterborne oocyst concentrations and consequent exposure risk are therefore very important factors in this study.

Before September 2007, the oocyst detection rate in final water at Clatto WTW (Dundee) was 4.4×10^{-4} per 10 litres compared with 13.2×10^{-4} per 10 litres for the Milngavie WTW (Glasgow) final water; hence consumers of Milngavie WTW (pre-September 2007) were exposed to water with 3 times the oocyst 'load' and were therefore likely to be at greater risk of exposure (table 1).

In the period after the introduction of filtration at Milngavie WTW, the oocyst count rate in final water from Loch Katrine decreased to zero (no oocysts detected), representing a complete removal of the waterborne oocyst exposure risk. In comparison, consumers of Clatto WTW final water were exposed to an average of 1.15×10^{-4} per 10 litres; this itself was a fourfold reduction in average waterborne oocyst exposure for Dundee study participants (down from 4.4×10^{-4} per 10 litres).

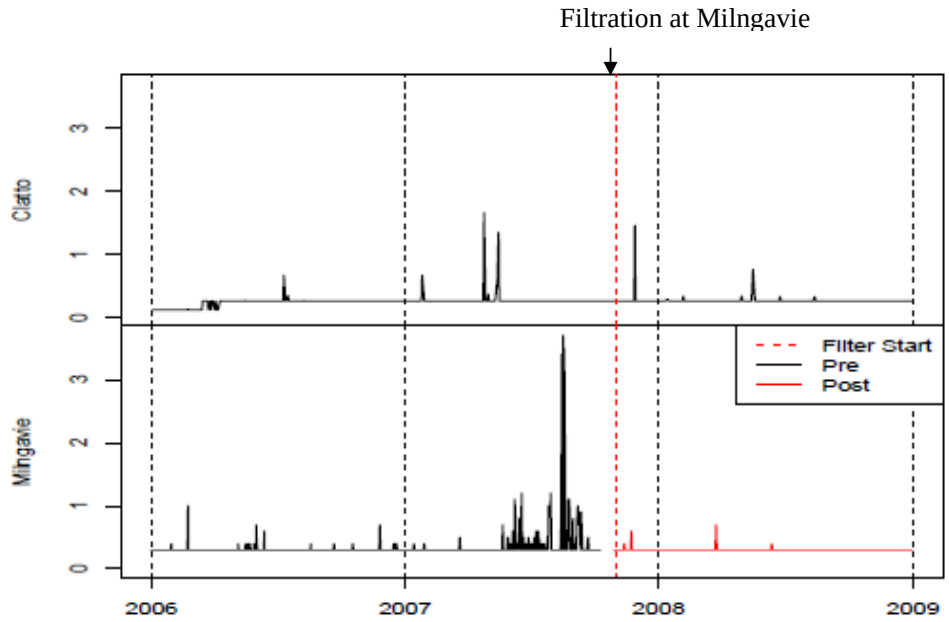
Over the study period, it is worth noting that the efficiency of removal of oocysts from raw water at Clatto, after September 2007, was comparable to that for the new WTW at Milngavie; 94% reduction vs 100% reduction. The net consequence for the study is that for both study cohorts, Glasgow and Dundee, there appears to have been a reduction in the risk of drinking oocyst-contaminated drinking water but the reduction in exposure risk for Glasgow blood donors was proportionately far greater. This finding for the Glasgow donors was expected. However, the finding for their Dundee counterparts was not. The reduced exposure of Clatto water consumers did not undermine the fundamental design of the study

nor the validity of the results, in that the 'power' of the study to detect a true difference was calculated on the seroprevalence rates in the human subjects and not based on the raw water oocyst concentrations.

3.2 Turbidity

Turbidity is a proxy measurement for water quality and provides a measure of the concentration of suspended particulate (solid) matter. It can be defined as cloudiness or opacity in the appearance of a water caused by solids, particles and other pollutants. The presence of turbidity in water supplies may be associated with rain run-off into reservoirs, carrying soil particles, animal faecal material and potentially micro-organisms. This in turn may indicate that increased water turbidity may be associated with an increased likelihood of pathogens such as *Cryptosporidium* oocysts in water. However, the correlation is not exact, as evidenced by previous outbreaks [32].

Figure 6. Turbidity measurements at Clatto and Milngavie WTW, 2006-2009.



The turbidity measurements in both supplies were generally low over the study period (figure 6). Turbidity rarely fluctuated above 0.25 units for Clatto and 0.3 for Milngavie. However, there was much more variability for the Milngavie WTW supply, with considerable spikes of turbidity before introduction of filtration. The peak values of turbidity were recorded from July 22nd to 23rd September 2007. The increase in turbidity at Milngavie over this period may have been associated with the mechanical disruption and building work in progress during the installation of the filtration system. The introduction of filtration at Milngavie WTW appears to have been temporarily associated with a reduced level of turbidity in the final water and less variability in the turbidity measurements. Turbidity levels are of interest as a measure of raw water quality, which subsequently impacts on the efficiency of filtration-based water treatment. However, the unreliability of correlations between measures of turbidity and oocyst counts mean that for this study, turbidity data is not particularly useful and is therefore not analysed further.

The presentation of the data and results of the statistical analysis is complicated by the nature and prolonged period of the study. The following section begins by presenting the data on the study participants and how the cohort changed over the period of study.

3.3 Donor attrition rates

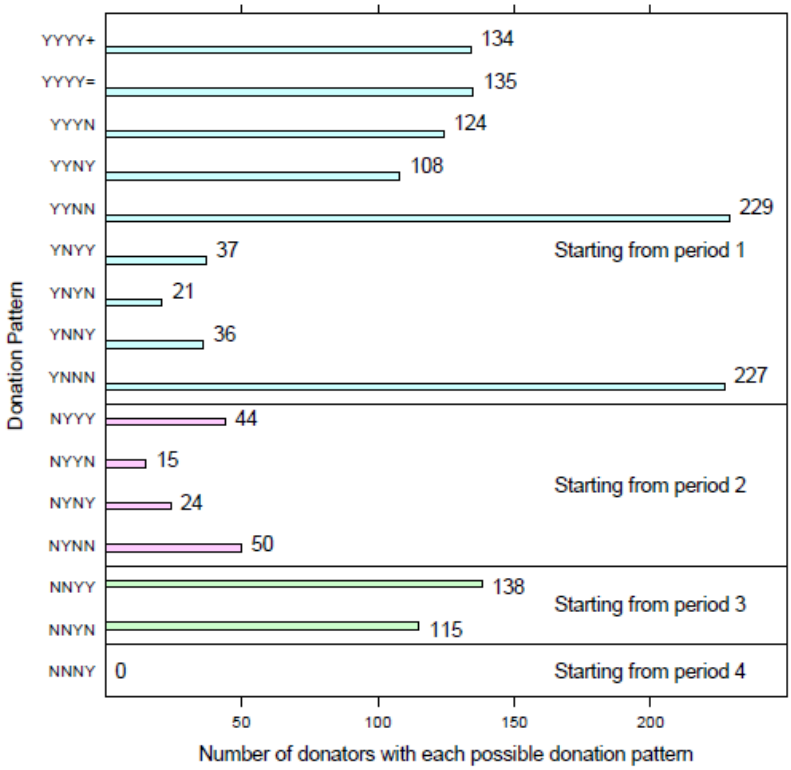
The data is derived from 3706 individual blood donation episodes, collected over four time periods, slightly less than the planned 4000. Each time a participant provided a sample, they were asked to complete another copy of the questionnaire. Participants were assigned a unique ID number, which was recorded for all future questionnaires that were completed to enable linkage. A small number of donors did not provide any questionnaire data (only a blood sample was taken) or their questionnaires were lost in transit, resulting in 54 blood samples having no corresponding questionnaire data. These were therefore

excluded from the analysis. Thus, 3652 blood samples with matching questionnaires over the four time periods, were analysed.

The total number of individual study participants over the whole period was 1437, greater than the planned 1000. In ideal circumstances, each donor would have donated in every time period however, it was accepted that this was not a realistic expectation over a 3-year study period and that attrition would occur. This is because not all participants continued from start to finish; some dropped out and were replaced by newly recruited donors. As such, this original concept of each participant providing one sample per period, proved impossible to achieve. The original cohort consisted of 791 donors from Glasgow and 260 from Dundee Blood Transfusion Service. By the end of period 4, there were 452 (of 791) original participants remaining from Glasgow and 213 (of 260) original participants from Dundee.

Some participants did not donate blood in all collection periods, while others donated more than once in some collection periods. The number of participants who provided samples over the study is shown in figure 7. It is worth noting that only 269 people donated at least once in all four periods (both cities). This emphasises the difficulties of retaining volunteer participation in this type of study. Although this is a small number, this group represents the 'core' cohort, unaffected by donor drop-out. It was hoped that provided the risk factor profile of this core group did not change over the 3-year period, it would provide the most valid indication of whether the enhanced physical treatment at Milngavie WTW was associated with changes in antibodies to *Cryptosporidium* among participants (provided the analysis was not compromised by the reduced numbers).

Figure 7. The number of donors by sample collection pattern - 'YYYY' denotes those that donated exactly once in each period. 'YYYY+' denotes those that donated more than four times i.e. more than once in one or more periods..



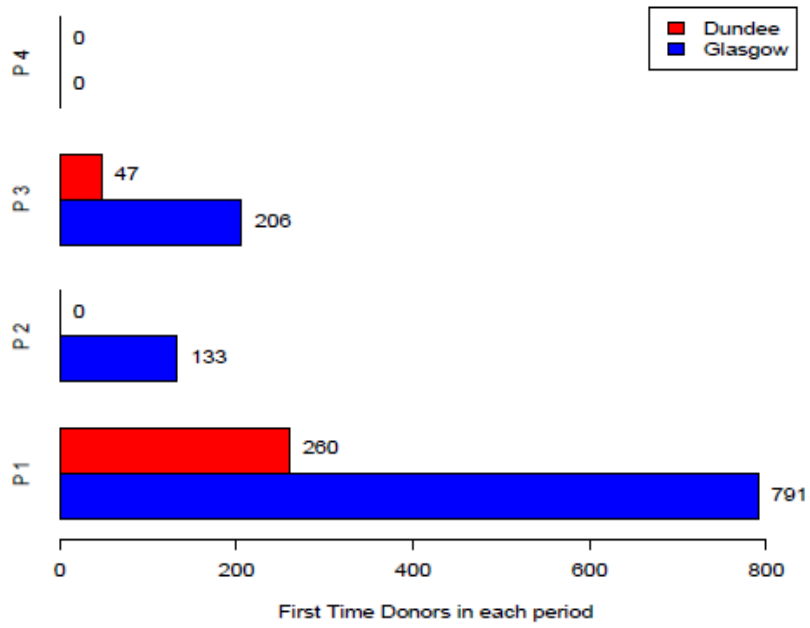
The number of original participants and new (replacement) donors, by city, over the four time periods is also given in table 2 and figure 8. In the second and third collection periods, there was a larger proportion of new recruits to the Glasgow cohort, due to the higher drop-out rate among the Glasgow donors. It is therefore important to assess how donor attrition and new recruitment of replacement participants affected the pattern of demographic, risk factors and serologic responses.

Table 2. Participants by study period; previous vs new recruits.

	Period 1	Period 2	Period 3	Period 4
Total participants per period	1051	883	861	911
Previous recruits*	1051	750	608	911
New recruits	0	133	253	0
Participants by city				
Glasgow	791	671	638	698
Dundee	260	212	223	213

* defined by someone who has donated once since period 1.

Figure 8. The number of new donors recruited within each collection period by donor city. No new donors were recruited in period 4.



3.4 Demographic comparison of Dundee vs Glasgow donors – period 1

Chi-squared tests were used to analyse demographic and risk factor differences between Glasgow and Dundee cohorts for the first collection period. The first collection period will provide the baseline data for other comparisons.

There were several statistically significant differences between blood donors in Glasgow and Dundee cohorts for the 1st collection period. Participants from Glasgow reported drinking more bottled water ($p< 0.024$) and less unboiled tap water than their Dundee counterparts ($p< 0.001$). While this is an interesting observation at the univariate level, it should not be over-interpreted; in view of the multiple testing and with Bonferroni adjustment (a statistical term to allow for the effects of multiple testing) only the unboiled tap water comparison is significant. Its significance will be explored more fully in the regression modelling.

3.5 Demographic of risk factor comparison of Glasgow cohort pre vs post-filtration

Because of donor attrition and the need to replace lost donors, the composition of both the Glasgow and Dundee cohort could have changed during the study. The first stage of comparison was to identify any significant difference in the Glasgow donor cohort composition between periods 1 and 2 versus periods 3 and 4.

Differences in questionnaire responses among the Glasgow donors before (periods 1 and 2) and after filtration (periods 3 and 4) were assessed to identify any significant differences. Such analyses should reveal if there were changes in demographic and/or risk factors before and after the change in treatment at Milngavie WTW, which could account for any detected difference in donor serological responses, in the Glasgow cohort.

Only 46% of all the Glasgow donors donated in either one of the first two periods (periods 1 and 2) and in one of the last two periods (period 3 and 4). This proportion of donors was not considered large enough to be representative of the entire Glasgow donor

cohort. Instead we assessed those donors that donated in either one of the first two periods and compared them to those that donated in either one of the last two periods. Thus, we considered responses from 924 participants for the group before filtration and 721 participants for the group after filtration. While the 924 participants pre-filtration are unique, it is acknowledged that some are also included in the second period and this dependency is not accounted for in the descriptive analysis.

Examples of the breakdown of epidemiological and risk factors for oocyst exposure for the Glasgow and Dundee cohorts are shown in tables 3 to 12.

Tables 3 and 4. Age of Glasgow cohort vs age of Dundee cohort, before and after filtration.

GLA

Age	**Group**		
	Pre	Post	Sum
17-44	592	387	979
45-64	318	308	626
65+	14	26	40
NA	0	0	0
Sum	924	721	1645

DUN

Age	**Group**		
	Pre	Post	Sum
17-44	134	92	226
45-64	123	110	233
65+	3	8	11
NA	0	0	0
Sum	260	210	470

Tables 5 and 6. Gender of Glasgow cohort vs Dundee cohort, before and after filtration.

GLA

Sex	**Group**		
	Pre	Post	Sum
Male	489	410	899
Female	435	311	746
NA	0	0	0
Sum	924	721	1645

DUN

Sex	**Group**		
	Pre	Post	Sum
Male	145	115	260
Female	115	95	210
NA	0	0	0
Sum	260	210	470

Tables 7 and 8. Children less than 5 in the household, by Glasgow and Dundee cohort, before and after filtration.

GLA

Have Kids < 5	Group		
	Pre	Post	Sum
Yes	62	50	112
No	859	668	1527
NA	3	3	6
Sum	924	721	1645

DUN

Have Kids < 5	Group		
	Pre	Post	Sum
Yes	18	17	35
No	242	193	435
NA	0	0	0
Sum	260	210	470

Tables 9 and 10. Pastime of swimming, by Glasgow and Dundee cohort, before and after filtration.

GLA

Swimming in UK	Group		
	Pre	Post	Sum
Yes	472	323	795
No	432	387	819
NA	20	11	31
Sum	924	721	1645

DUN

Swimming in UK	Group		
	Pre	Post	Sum
Yes	142	91	233
No	118	118	236
NA	0	1	1
Sum	260	210	470

Tables 11 and 12. Ownership of pets with diarrhoea, by Glasgow and Dundee cohort, before and after filtration.

GLA

Pets had diarr.	Group		
	Pre	Post	Sum
Yes	6	4	10
No	246	195	441
NA	25	3	28
Sum	277	202	479

DUN

Pets had diarr.	Group		
	Pre	Post	Sum
Yes	1	1	2
No	96	76	172
NA	9	2	11
Sum	106	79	185

There were several statistically significant differences between participants from Glasgow who provided samples in the pre and post-filtration periods (periods 3 and 4). Participants from Glasgow in the post-filtration period were more likely to have pets that had a past episode of diarrhoea within the relevant period (tables 11 and 12 (p=0.002)), more likely to have had contact with exotic and/or farm animals (p=0.048) but were less likely to report swimming (p=0.011) than those participants from Glasgow who provided samples before the filtration system was introduced (periods 1 and 2). With Bonferroni adjustment, only having pets with diarrhoea remains statistically significant. However, even then numerically they represented a tiny fraction of the relevant donor population and hence having a pet with a history of diarrhoea is not of material importance as a confounding factor in the study.

3.6 Demographic comparison of Dundee cohort pre vs post-filtration

In this comparison, data was assessed to determine if there were any significant differences in the Dundee participants between periods 1 and 2 and periods 3 and 4. Such analyses should reveal if there were demographic and/or risk factors which changed in profile before and after September 2007 and which could be associated with any detectable difference in serological responses observed in the Dundee cohort.

As with Glasgow participants, only 53% of donors from Dundee donated in either one of the first two periods before filtration and either one of the last two periods after filtration. This proportion of donors was not considered to be representative of the whole Dundee cohort. Instead, an analysis was carried out on those that only donated in one of the two groups i.e. those that only donated in one of the first two periods and those that only donated in the last two periods. Thus, we considered responses from all 307 Dundee participants, resulting in 260 responses for the pre-filtration group (periods 1 and 2) and

210 for the post-filtration cohort (periods 3 and 4). Some (63) participants appeared in both groups having donated both before and after September 2007.

There were few significant differences in demographics and expopsure factor variables between the pre-filtration and post-filtration participants from Dundee. As with the Glasgow cohort, there were some differences which can be explained by the progression of time (for time/duration-related variables). The most notable difference between Dundee participants before and after September 2007 was that those people who donated after (periods 3 or 4), were less likely to report swimming in the previous 12 months ($p=0.015$, not significant after Bonferroni adjustment). This finding is in accordance with findings for the Glasgow cohorts, before and after filtration.

3.7 Discussion of demographic and exposure factor comparisons between citiy donor populations

Having identified that there were some differences within the two cohorts (over the study periods in total), it was important to assess Glasgow and Dundee participants for demographic and 'risk factor' profiles throughout the course of the study, in order to ascertain if any differences in serological responses were directly associated with the introduction of enhanced filtration at Loch Katrine WTW in September 2007. It was important to be aware of any such effects, in case they might influence the final conclusions about the impact of enhancing filtration at Loch Katrine. Sections 3.4-3.6 repeated the analysis for the two populations. Although it is interesting to note that Glasgow donors consumed more bottled water and less unboiled drinking water than their Dundee counterparts in period 1, this observation was a univariate measurement and its significance was not known until the mixed modelling was performed. The only significant finding that Glasgow donors drink less unboiled tap water than Dundee counterparts may be due to local factors e.g. the taste of water prior to filtration, the residual impact of

warnings issued for previous incidents with “boil water notices” in Glasgow or other unknown factors. The significance of this finding is explored further in the more detailed analysis later.

Donors from both Glasgow and Dundee were less likely to report swimming after filtration was introduced (periods 3 and 4) than before. Reasons for this are unclear but may be multi-factorial. One potential factor might have been the information leaflet, given out to blood donors during recruitment. This highlighted swimming as a potential risk factor for acquiring cryptosporidiosis and might possibly have discouraged some people from swimming. This is speculative and cannot be verified. The change may have been a reflection of wider population trends and factors.

In summary, there were very few significant differences between Glasgow and Dundee donors in terms of demographics or recognised ‘risk factors’ for exposure to *Cryptosporidium*. The relative lack of significant basic differences in demographic and risk factor profiles between Glasgow and Dundee cohorts is encouraging and provides confidence in the further analysis to be performed on the two cohorts. However, the results presented in sections 3.4-3.6 rely on each observation being independent of the others, which is not actually the case. Assessment of a single, dependent variable can be statistically analysed using univariate techniques e.g. by chi-squared analysis. More complex multivariate analysis is therefore required to address these inter-dependencies and provide more robust evidence of significant differences, if any really exist. This truly elucidates which parameters are the most important in terms of serological results and how these differ over time and by city donor population. This is described further using mixed modelling techniques.

3.8 Descriptive analysis of serological responses pre vs post-filtration

Of the 3706 separate questionnaires from the study participants, 25 participants had no matching blood sample. Therefore, these were excluded from this analysis. Serological responses detected in samples from participants are expressed in terms related to the 'positive' controls, as discussed in the methods section. A total of 3681 samples from all participants in both cohorts were analysed. However, there were 19 blood samples donated 'anonymously' where no related questionnaire was completed or could be traced. Hence, the total number of samples with matching questionnaires was actually 3662.

The main measure of seropositivity (the level of *Cryptosporidium* specific antibody response) used is the 'percentage positive response' or PPR. This is a relative and not absolute measure. The PPR is relative to the positive rabbit sera control is plotted for all sample results (figure 9). This histogram shows that the distribution of serological responses was skewed, with most samples showing less than 60% 'positivity'. Such a skewed distribution creates analytical problems, hence to reduce the effect of the skew, the data was transformed logarithmically (after adding 1 to all values to make them positive). The 'normalised' data are then more suitable for appropriate statistical analyses. A histogram of the log transformed PPR values is shown in figure 10; the spike at 0 corresponds to those samples with no detectable antibodies. The summary statistics in the following sections are calculated based on analysis of this transformed data and are used to identify relative differences across collection periods and participant city cohorts.

Figure 9. Distribution of percentage positive responses (PPR) (n= 3681)

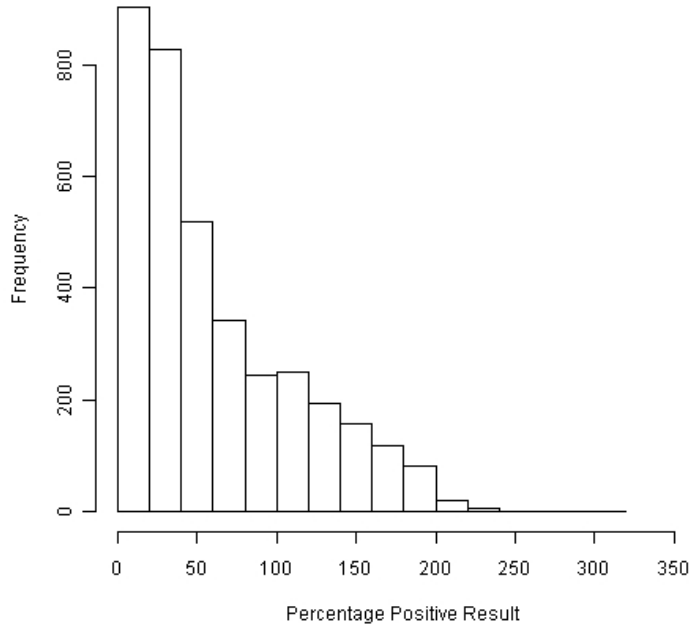
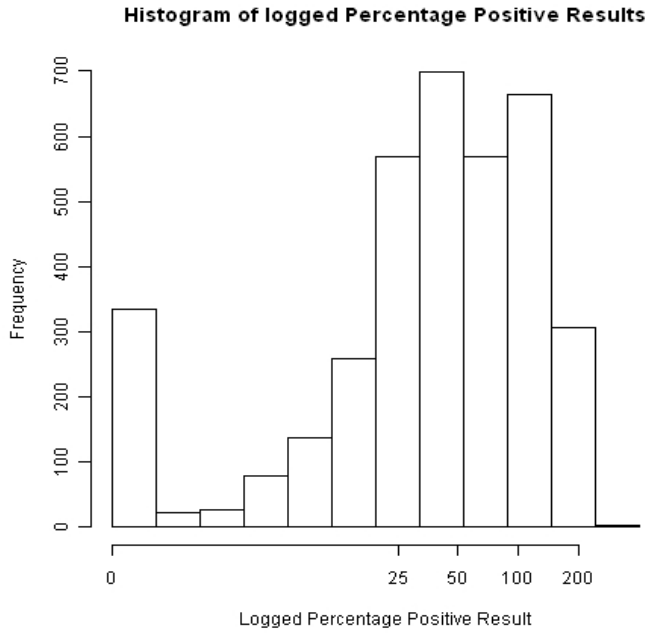


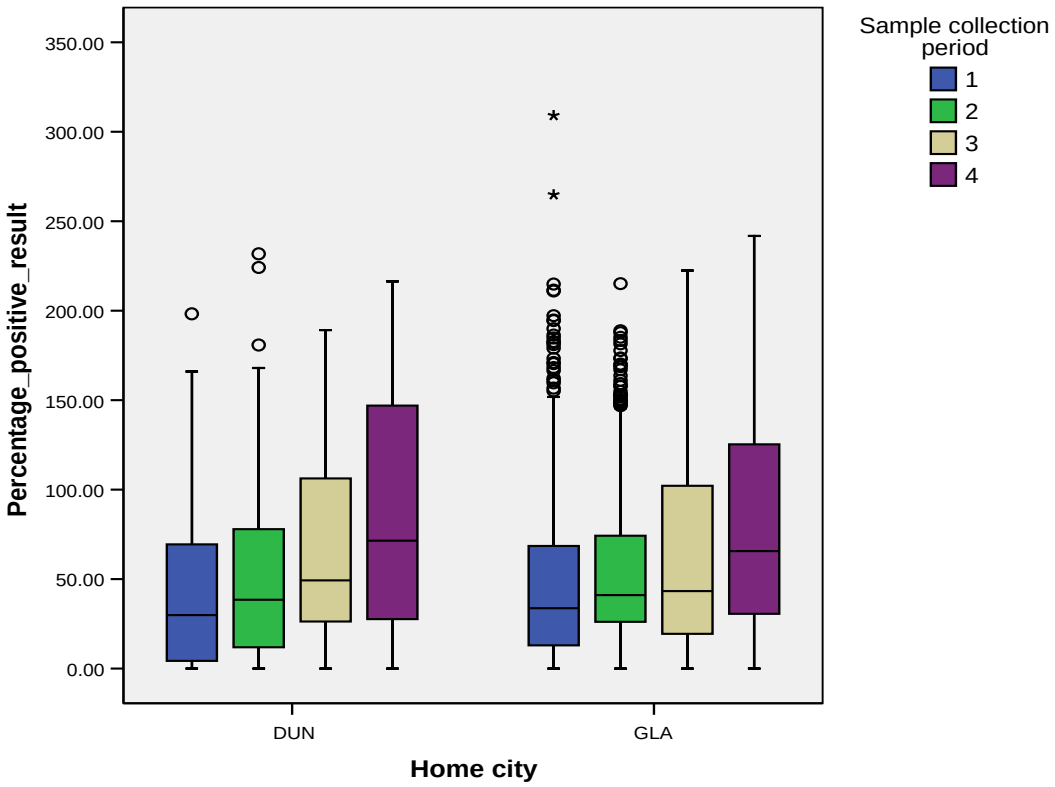
Figure 10. Distribution of log transformed percentage positive response (PPR) (n=3681).



The median and quartile raw PPR data (+/- the standard deviation) were calculated for donors by city and collection period and presented as a box and whiskers plot (figure 11). The figure illustrates a consistent rising trend in serological responses to *Cryptosporidium* in both cohorts over the collection periods. However, between collection periods 2 and 3 (when enhanced filtration was introduced to Loch Katrine), there is very little increase in serological responses in the Glasgow cohort compared with Dundee. The plateauing (coinciding with the introduction of filtration) was short-lived however as PPR for both Glasgow and Dundee cohorts was similar in collection period 4. The period 3 flattening of the trend for an increase in PPR over time, was therefore a relatively temporary phenomenon.

The median and quartile (log) PPR value (+/- the standard deviation) were calculated for donors by city and by collection period and presented as a box and whiskers plot (figure 13). The figure is very similar to the figure 11 but these results should be interpreted with caution as the repeated observations on donors and the introduction of new donors is not taken into account. All the results presented in this section using univariate analysis rely on an assumption that all the observations are independent of each other in that some of the same participants provided samples in multiples periods. Mixed modelling is required to more accurately evaluate any significant differences allowing for the non-independence of individual results.

Figure 11. Box and whiskers plot of raw percentage positive response (PPR) by city/period (sequential plots illustrate the chronological collection periods).



From figure 11, a number of outlying results (indicated by asterisks/noughts) are evident for both populations, hinting at strong antibody responses to *Cryptosporidium*. Perhaps the results are not entirely surprising in that it would be unusual for many people to have been infected with *Cryptosporidium* in the general population. The boxplot can be interpreted in figure 12:

Figure 12. Interpreting a boxplot.

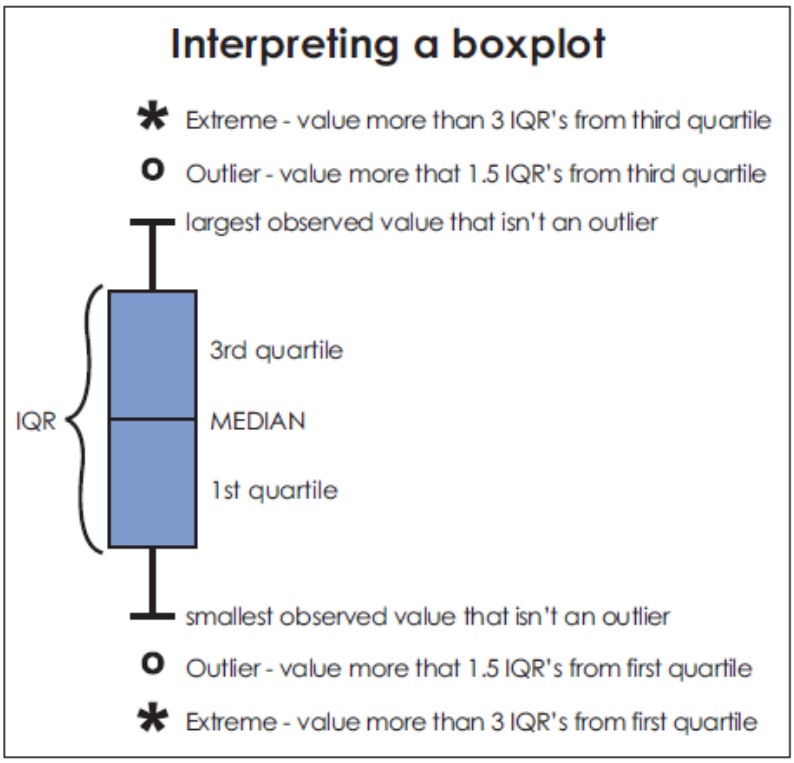
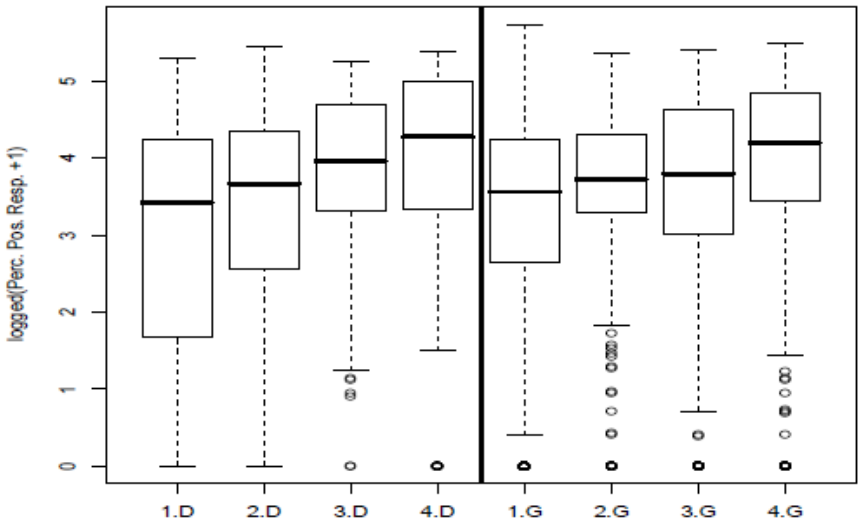


Figure 13. Box and whiskers plot of log percentage positive response by city/period (1.D denotes 1st collection period in Dundee; 1.G denotes 1st collection period in Glasgow and so on.)



3.9 Serological responses (PPR) which are zero

In figure 10, the number of serological responses which are zero is relatively high. This distorts the normal distribution 'bell-shaped' curve; ideally for a mixed model analysis, the data would conform more closely to a normal distribution. Consideration was given to excluding the zero responses from the analysis, however this changed the initial modelling results. This suggests that the zero PPR were not 'balanced' i.e. there are either more serological responses that are zero in one cohort compared with the other, or there are more serological responses that were zero in one collection period compared with another. Zero responses were therefore considered separately in order to assess if they would introduce bias into the mixed modelling analysis.

There were 325 serological responses with a PPR value of zero. Table 13a shows that there was a relatively higher proportion of 'zero'serological PPR from the Dundee cohort compared with Glasgow. This equated to 92.9% of Glasgow donor samples with a PPR>0, compared with 88.7% of Dundee donor samples (table 13b). This was supported by a chi-squared test of independence giving a highly significant 'p' value (< 0.001). One possible suggestion might be that as the Clatto supply, which provides Dundee residents has been filtered for many years. Hence, it is possible that the Dundee cohort were less exposed to *Cryptosporidium* via drinking water than those in Glasgow. This is consistent with the hypothesis of the study. However, other unknown historical factors may also be involved.

Table 13a. Comparison of proportions of all PPR>0 compared to proportions of PPR=0.

	PPR>0		PPR=0	
Cohort	**n**	**%**	**n**	**%**
Glasgow	2777	76	212	65
Dundee	885	24	113	35
Total	3662	100	325	100

Table 13b. Comparison of cohorts where PPR=0 or is >0.

	PPR>0	**PPR=0**	**Totals**
Cohort			
Glasgow	2777 (92.9%)	212 (7.1%)	2989
Dundee	885 (88.7%)	113 (11.3%)	998
Total	3662	325	3987

The zero responses were also analysed by distribution over the collection periods (Table 14).

Table 14. Comparison of PPR (zero and >0) by collection period.

	Dundee						**Glasgow**						
	>0		Zero		Total		>0		Zero		Total		
Period	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**Total**
1	208	80	53	20	261	24	730	89	94	11	824	26	1085
2	160	78	45	22	205	23	654	97	18	3	672	77	877
3	227	99	2	1	229	26	582	91	61	9	643	74	872
4	177	93	13	7	190	23	599	94	39	6	638	77	828
Total	772	87	113	13	885	24	2565	92	212	8	2777	76	3662

The table clearly shows that for both Dundee and Glasgow combined, there are more serological responses which are zero in the first collection period compared with the other collection periods (33% vs 25% for 2nd period). This is supported by a chi-squared test with a significant p value ($p< 0.001$).

We investigated the risk profile of those who had detectable serological responses and those who had zero serological responses and assessed differences in questionnaire responses on risk factor exposures using chi-squared tests (Table 15).

Table 15. Summary of chi-squared tests on serological response data; PPR zero vs non-zero.

Risk factor exposure	'p' value	Reason
Age	< 0.001	Younger people are more likely to have zero responses
Consumption of unboiled drinking water	0.266	Not significant
Consumption of unboiled drinking water (home)	0.311	Not significant
Consumption of unboiled drinking water (work)	0.487	Not significant
Consumption of bottled water	0.066	Not significant
Amount of bottled water consumed	0.534	Not significant

The only significant difference noted between participants who had a detected serological response and those who did not, was that the 'zero' responders were

significantly younger ($p < 0.001$). This is consistent with the theory that the risks of exposure to *Cryptosporidium* and hence production of a serological response is likely to increase with age.

3.10 Discussion of serological responses – descriptive analysis

It is important to evaluate any differences in the donor serological responses, between those from Glasgow and Dundee cohorts, to give us an approximation of how serological responses change over time and between the city cohorts. While the data shows that there is an underlying rising trend in PPR in both cities, it is apparent that between periods 2 and 3, there was little change in PPR for participants from Glasgow. This coincided with the timing of the introduction of enhanced filtration in the Loch Katrine supply in September 2007. This observation, based on the univariate analysis, suggests that filtration may have had a temporary impact on the PPR among the Glasgow cohort. However, as with previous univariate analysis, we need to interpret this with caution. The PPR results on sequential samples taken from the same donor are not independent measurements and so mixed modelling is required to appropriately test for differences associated with introduction of enhanced treatment at Loch Katrine WTW.

3.11 Logistic mixed modelling of (binary) serological responses

Having analysed the demographics, risk factors and serological responses using univariate analysis, we now use multi-variate analysis (models) to ascertain how such factors interact with each other and attempt to elucidate the most significant factors which affect serological responses over time.

Within the logistic models, we fit a series of contrasts (comparisons) which allow us to test various differences in serological responses linked with differences between the cohorts and collection periods (Table 16).

Table 16. Statistical 'contrasts' used in the binary mixed effect model.

Contrast	Analysis	+/- value indicates
City	Difference in responses between cohorts	Negative value implies fewer/lower serological responses in Glasgow
PrePost	Differences before and after filtration at Milngavie WTW	Positive implies more/higher responses after filtration
p2vp1	Differences between periods 2 and 1	Positive implies more/higher responses in period 2
p4vp3	Differences between periods 4 and 3	Positive implies more/higher responses in period 4
CityByPrePost	Different pre/post effect in Glasgow compared with Dundee	Negative implies fewer responses after filtration at Milngavie WTW compared with Dundee
CityByp2vp1	Different effect in period 2 versus period 1 for Glasgow	Positive implies Glasgow has more/higher responses compared with Dundee in period 2
CityByp4vp3	Different effect in period 4 versus period 3 for Glasgow	Positive implies Glasgow has more/higher responses compared with Dundee in period 4

A ‘contrast’ is a statistical term used to explain the relationship between each measurable parameter under assessment. It is an efficient way of measuring the important differences in the above table. There are two city cohorts and four different time periods – this results in eight main observations, assessing the proportion of people with a serological response and ultimately allowing 7 different comparisons – or contrasts. If all donors were measured in all 4 periods then these contrasts could be interpreted independently of each other. Some contrasts – *CityByPrePost, CityByp2vp1, CityByp4vp3* – are interactions and will be investigated first.

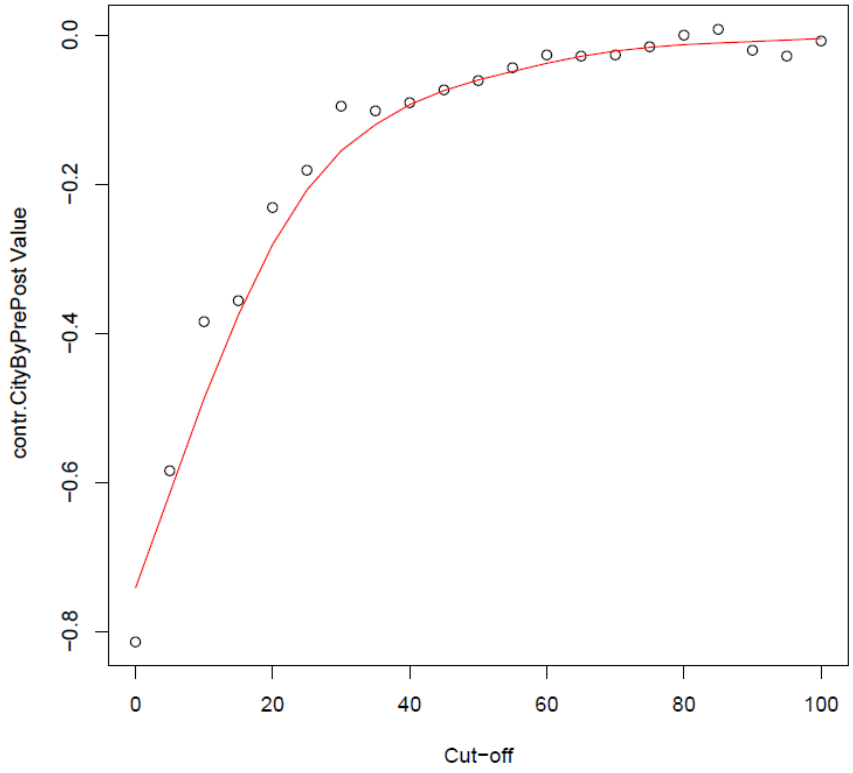
The main thrust of the study asks; is there an effect on serologic responses associated with the change in water treatment introduced at Loch Katrine in September 2007 and is there a significant difference between Glasgow and Dundee cohorts? Thus, the most important contrast is the *CityByPrePost*. This assesses if any change in PPR (in the post-filtration period compared to pre-filtration period) is the same for the Glasgow cohort as for Dundee. If it is the same, then the introduction of filtration in Glasgow has had no measurable effect. If this value is negative then there are lower serological responses after filtration in Glasgow compared with Dundee and this provides evidence that the introduction of filtration has had an impact in reducing exposure to *Cryptosporidium* oocysts through drinking water. If this value is positive then there are higher serological responses after filtration in Glasgow compared with Dundee; this implies that the introduction of filtration may have had a beneficial impact.

In practice, the interaction effects between *City* and *p2vp1* and *p4vp3* were minor (not statistically significant) so we excluded those from the model and do not consider them further.

Figure 14 shows the value of the contrast, *CityByPrePost* for different cut-offs of the PPR values (as explained in the methods section); in effect, this is a sensitivity analysis, which considers what would be deemed a ‘positive’ serologic responses vs a ‘negative’ serologic response. The line is a smoothed fit to these ‘cut-off’ values. The exact

statistical explanation for smoothing goes beyond the remit of this report but essentially allows for removal of variation/noise while being statistically robust. We have not included the graphs for all the contrasts listed in table 16 but use figure 14 as an example of how we selected appropriate cut-offs to assign a positive serological value for the serologic response (versus a negative value) for further analysis using the logistic mixed model. As *CityByPrePost* is the most important contrast, we used this to investigate the sensitivity of the analysis to the choice of cut-off denoting a positive response. This identified that the estimated effect is always negative (except at cut-offs approaching 100%); this indicates that there are fewer/lower serological responses in the Glasgow cohort samples after filtration compared with Dundee counterparts. This provides evidence that the introduction of filtration was associated with an impact in terms of apparently reducing exposure to *Cryptosporidium* oocysts.

Figure 14. Binary effects model with contrast *CityByPrePost*.



Logistic models permit data with 'binary' responses (yes/no, positive/negative) to be analysed using regression models. The PPR is a continuous variable (%) and is therefore not a 'binary' value so has to be converted into a 'binary' equivalent for the logistic model i.e. those who are 'positive' denotes participants who are deemed to have been exposed to *Cryptosporidium* and 'negative' for those who have not. We can choose different cut-off values above which someone is considered to have been exposed to the organism or its antigenic proteins i.e. the threshold for the PPR equated to 'exposure' can be calibrated to alter who is counted as having a 'positive' response. Responses below or equal to this calibrated value are considered to be negative i.e. have had no exposure to *Cryptosporidium*. To an extent, this is an arbitrary decision. There is no definitive or empirical method to set an appropriate 'cut-off' value. A 100% response simply means a response equivelent to that of an artificially exposed mouse, not a realistic human environmental exposure scenario.

Adopting a very strict standard (e.g. using the positive control as 100%), would mean that only participants who had a PPR of 100% or more would be counted as 'positive'. However, the participants' exposure to oocysts was undoubtedly low-level environmental exposure rather than the equivalent of being deliberately immunised (as was the rabbit control). This is borne out by approximately 20% of results being at or above a PPR of 100%. Adopting a 100% 'cut-off' for this analysis, would therefore exclude the majority of participants who had actually been exposed to oocysts. Adopting a 0% PPR 'cut-off' would result in the opposite extreme, which would only exclude participants who had no evidence of being exposed i.e. approximately 9% of participants had a PPR=0 i.e. 91% of the cohort would be considered positive.

In this analysis, we have therefore used various intermediate 'cut-offs' to look for evidence of a consistent effect between having a higher 'cut-off' and the calculated value for a 'contrast'. The final selection of which is an appropriate but also reasonable and

scientifically justifiable cut-off value, is based on an understanding of the science, the implications of low-level environmental exposure and precedent as cited in comparable *Cryptosporidium* serology studies. As an example, using Figure 13, if we select 'cut-offs' at 0, 10, 20 and 30 percent of the PPR, we are more likely to report a statistically significant difference between city donor cohorts, if one exists. This type of sensitivity analysis has been performed for other studies assessing seroprevalence [14, 18]. The number of positive and negative donors under each cut-off value is shown in Table 17:

Table 17. The binary splits of responses using different PPR cut-off values from 0-310% (n= 3667).

Cut-off	Negative	Positive	Percent Positive
0	325	3342	91.1
10	566	3101	84.6
20	900	2767	75.5
30	1323	2344	63.9
40	1730	1937	52.8
50	2025	1642	44.8
60	2250	1417	38.6
70	2435	1232	33.6
80	2592	1075	29.3
90	2710	957	26.1
100	2837	830	22.6
110	2971	696	19.0
120	3087	580	15.8
130	3191	476	13.0
140	3280	387	10.6
150	3380	287	7.8
160	3437	230	6.3
170	3496	171	4.7
180	3555	112	3.1
190	3609	58	1.6
200	3636	31	0.8
210	3644	23	0.6
220	3657	10	0.3
230	3661	6	0.2
240	3664	3	0.1
250	3665	2	0.1
260	3665	2	0.1
270	3666	1	0.0
280	3666	1	0.0
290	3666	1	0.0
300	3666	1	0.0
310	3667	0	0.0

Table 17 illustrates the effect of using different 'cut-offs' to designate if a participant had exposure to oocysts (positive) or not (negative). A 'cut-off' of 0 PPR means that 91.1% of participants are considered to have been exposed. Intuitively, this appears to be a high proportion of the population under study. Selecting a 'cut-off'of 100 PPR yields 22.6% of participant samples, which implies that under a quarter of the population under study had a response equivalent to that of the 'positive' rabbit sera control.

Given that our study population are predominantly people who have no clinical history of infection with *Cryptosporidium*, it would seem reasonable not to expect them to have as strong a serological response to *Cryptosporidium* oocyst proteins as an immunised rabbit or clinical human case. We are looking for evidence of historic exposure to oocysts at environmental levels, which have not been sufficient to cause a clinically confirmed illness due to *Cryptosporidium* in our participants. Hence on that basis, it is reasonable to set a 'cut-off' for the response (as PPR), which is likely to indicate past environmental exposure, not recent clinical illness. An upper limit of 50 PPR was therefore selected for the 'cut-off' options in the following analysis on the basis that it was unreasonable to expect low-level environmental exposure to result in a serological response, which persists above 50% of that likely in a clinically-infected case [14]. The justification for this is covered further in the final discussion. The sensitivity analysis therefore only assesses further PPR 'cut-off' values from 0 to 50%.

Using each 'cut-off', the same logistic mixed effects model is then fitted. To test for differences in serological responses between Glasgow and Dundee cohorts, and in time over different collection periods, a series of contrasts is utilised. The fitted values of the contrasts under each logistic model are shown in Table 18.

Table 18. The percent differences implied by the contrasts calculated under each logistic model which was fitted to a different 'cut-off'.

Contrast	Cut-off at 0 %			Cut-off at 10 %			Cut-off at 20 %		
	%Diff	P-Value	Sig?	%Diff	P-Value	Sig?	%Diff	P-Value	Sig?
City	4	0.868	N	34	0.036	Y	12	0.342	N
prePost	242	<0.001	Y	102	<0.001	Y	82	<0.001	Y
p2Vp1	118	<0.001	Y	122	<0.001	Y	102	<0.001	Y
p4Vp3	−70	<0.001	Y	6	0.658	N	36	0.016	Y
CityByPrePost	−126	<0.001	Y	−66	<0.001	Y	−44	<0.001	Y
CityByp2Vp1	136	<0.001	Y	52	<0.001	Y	44	<0.001	Y
CityByp4Vp3	236	<0.001	Y	64	<0.001	Y	30	0.037	Y

Contrast	Cut-off at 30 %			Cut-off at 40 %			Cut-off at 50 %		
	%Diff	P-Value	Sig?	%Diff	P-Value	Sig?	%Diff	P-Value	Sig?
City	6	0.559	N	0	0.948	N	−8	0.464	N
prePost	78	<0.001	Y	92	<0.001	Y	98	<0.001	Y
p2Vp1	56	<0.001	Y	34	0.003	Y	24	0.041	Y
p4Vp3	46	<0.001	Y	74	<0.001	Y	80	<0.001	Y
CityByPrePost	−16	0.035	Y	−16	0.036	Y	−12	0.142	N
CityByp2Vp1	2	0.807	N	−2	0.818	N	4	0.678	N
CityByp4Vp3	24	0.063	N	10	0.434	N	2	0.911	N

In table 18, the column headed '%Diff' is calculated from the fitted logistic mixed effects model, which predicts the log 'odds ratio' of a serological response. The odds ratio is the probability of a response divided by the probability of a nil response; it is a measure of the likelihood of a response relative to no response. An odds of 1 means that a positive response is as likely as a negative response and an odds ratio of greater than 1 means that a positive response is more likely than a negative one. The odds ratio is the ratio of the odds for one group compared to another group; for example the Dundee cohort compared to Glasgow. If the odds ratio is 1 then the odds of a positive response are the same in the two groups. If it is greater than 1 then there is a greater chance of a positive response in the Dundee group compared with Glasgow; if it is less than one then there is a lower chance of a positive response in the first group compared with the second. Odds ratios are often expressed as percentage differences. If an odds ratio is 1.5 then the odds of a positive

response in the first group is 50% greater than in the second; if it is 2 we say the odds are 100% greater (or doubled). The column headed '%Diff' in Table 18 is therefore a percentage difference in odds ratios associated with the contrast.

Pre-post contrast

This analysis looks for effects only attributable to differences in PPR associated with the collection periods P1+P2 (pre-filtration at Milngavie WTW) and P3+P4 (post-filtration at Milngavie WTW) for the entire participant cohort taken together (Dundee and Glasgow). It does not separately analyse any effects associated with the individual city cohort; that follows in later models. Table 18 shows that the *prePost* contrast is positive and statistically significant for all the models. This suggests that the proportion of positive serological responses to *Cryptosporidium* is consistently higher in the post-filtration period compared to the proportion in the pre-filtration period for all the participants, in both cohorts. The large percentage change (from 98 to 242%) for the *prePost* contrast suggests that the change in proportion of positive serological responses is most strongly affected by the difference between the pre- and post-filtration levels of serological response. At a cut off of 30%, the estimated effect is 78%. As there is an interaction effect involving *prePost* this estimate refers to the reference level for the interacting term, *City*, which is Dundee. In Dundee, therefore the odds of a positive serological response is 78% greater in the post filtration periods compared to the pre-filtration periods.

P2vP1 contrast

The *p2vp1* contrast assesses effects associated with the change from period 1 to period 2 only, for the whole participant cohort. The *p2vp1* contrasts are positive and statistically significant in all models therefore confirming that there is evidence to suggest that the proportion of positive serological responses in period 2 is higher than in period 1

(see Figure 10). This is consistent across all models but declines in strength as the 'cut-off' value increases from 0 to 50%.

P4vP3 contrast

The *p4vp3* contrast assesses effects associated with the change from period 3 to period 4 only, for the whole participant cohort. The *p4vp3* contrast is somewhat harder to interpret; as the 'cut-off' increases, the percentage difference for this contrast also increases (from -70% Difference for PPR=0 to 80% Difference for PPR=50 (Table 18). However, this increase becomes smaller between subsequent cut-offs suggesting it becomes non-significant (40 and 50% cut-off models). The evidence from the *Prepost*, *p2vp1* and *p4vp3* analyses supports the hypothesis that there is a general increasing trend in positive serological responses over time and this is in agreement with the earlier descriptive analysis (section 3.8).

City contrast

The city contrast looks exclusively for evidence of differences associated with the individual city cohorts irrespective of time period differences. The *City* contrast suggests little difference between the cities independently. A positive percentage difference suggests a greater proportion of positive serological responses in the Glasgow cohort. However, these differences are quite small numerically and are only large enough to be statistically significant in the 10% cut-off model. The finding is not consistently significant across the range of 'cut-off' values and is therefore not a sustained difference, implying it is unlikely to be important. As *City* is involved in three interaction effects, it is not easy to interpret. However at a cut-off of 30%, the interactions with *CityByp2vp1* and *CityByp4vp3* are not statistically significant and will be ignored. This just leaves the *CitybyPrePost* interaction and the estimated effect of *City* refers to the lower level of *PrePost*, namely pre-filtration. Thus *City* is a comparison of Glasgow relative to Dundee

in the pre-filtration period and the estimated effect is 6%, implying that the odds of a positive serological response (at 30% cut-off) is 6% higher (not significant) in Glasgow compared to Dundee.

We next consider the interaction contrasts *CityByPrePost*, *CityByp2vp1* and *CityByp4vp3* to assess if there is a difference in the effects associated with duration between Dundee and Glasgow.

CityByPrePost contrast

CityByPrePost assesses any differences before (periods 1 and 2) and after (periods 3 and 4) filtration in the Glasgow cohort compared with Dundee. By the structure of the *CityByPrePost* contrast, a negative percentage change will indicate that the *prePost* effect is smaller in Glasgow. The consistently negative contrast values for this section implies that the increase in seropositivity over time is consistently less in Glasgow participants. The increase in the proportion of positive serological responses in Glasgow participants between the pre and post-filtration period is smaller than any corresponding increase observed in the Dundee cohort. While the magnitude of the difference between Dundee and Glasgow decreases as the cut-off increases, it is statistically significant in all models considered up to analysis at the 50% 'cut-off' level. This suggests that there is an interaction between the water source to which participants were exposed but that the magnitude of this effect is less pronounced as more participants with stronger serological responses (>50 PPR) are included. This could be consistent with a weak immunogenic effect of a low-level exposure (e.g. drinking water) being less of an influence on the host immune response to *Cryptosporidium*, than other environmental stimuli e.g. exposure via animals or contact with infected cases.

At a cut-off of 30%, the estimated effect is -16%. This means that the odds ratio of a positive serological response in Glasgow donors (post-filtration compared to pre-filtration) is 16% less than the same odds ratio for Dundee donors. As above, when

discussing the *PrePost* contrast the odds of a positive serological response is greater in the post-filtration period compared to the pre-filtration period. The 16% reduction in Glasgow means that the increase in Glasgow was smaller than the increase in Dundee.

CityByP2vP1 contrast

This contrast looks at evidence of differences between the city cohorts between periods 1 and 2. The PPR change for *CityByp2vp1* indicates that the increase in proportion of positive serological responses between collection periods 1 and 2 is larger in Glasgow. This effect decreases in magnitude as the PPR 'cut-off' increases and stops being statistically significant when the cut-off is > 20%. This suggests that since the effect does not persist across the range of PPR 'cut-off', that it is a relatively weak and unimportant effect.

CityByP4vP3 contrast

As before, this contrast assesses evidence of a difference between the city cohorts only between periods 3 and 4. This contrast also shows that there is a larger, relative increase in positivity among Glasgow participants but generally the effect is larger i.e. the increase between collection periods 3 and 4 is larger in Glasgow than in Dundee (and larger than it was between the two cohorts in periods 1 and 2).

To summarise, overall the trend for an increase in the proportion of positive serological responses over time, is smaller in the Glasgow cohort compared with Dundee. However, looking at differences within pre/post periods provides more complicated patterns of variation. It can be concluded that the logistic mixed models provide statistically significant evidence that there is some association between the introduction of the new filtration system at Milngavie WTW, with a reduction in the rate of increase of seropositivity over time in Glasgow by approximately 16-44% (according to 'cut-off' over the range 20-40%). However, this effect was apparently a temporary effect that did not

persist beyond period 3; the proportion of 'positive' serological responses (across the spectrum of 'cut-offs' used to denote 'positive' from 0 PPR to 50 PPR) increased in Glasgow by a similar amount to Dundee from period three to four. The net effect appears to have been a 'step-change' in the seropositivity profile of Glasgow study participants. The magnitude of that step-change may give a preliminary indication of the magnitude of the effect of low-level exposure to *Cryptosporidium* via drinking water from Loch Katrine, before filtration was introduced.

3.12 Linear mixed modelling of serological responses

To further assess for differences in serological responses related to city and collection period, a linear mixed effect model was used. In contrast to the logistic mixed model, the linear model is fitted directly to the PPR serological results, which gives greater power to detect differences. Linear models should only be used on data that is approximately normally distributed. Hence, as illustrated earlier, the distribution of PPR is not normal. In order to satisfy this condition, the model was fitted to a transformed version of the serological response data. As before, a series of contrasts were used to test for particular differences within the data. Interpretation of the model is complex and therefore a two-step process is used, in which two models were employed:

1. the first model uses the square root of the donor serological response. From this, significant contrasts are determined.
2. the second model assesses log of the donor serological response, only including those contrasts which were significant using the first linear model above.

Diagnostic plots (figure 15) indicate a linear fit of the model. Figure 14 is used to validate that the data used for this model conforms approximately to a 'normal' distribution using the square-root transformation. The coefficients of the model are given in Table 19.

Figure 15. Linear model of serological responses. The approximately 'straight' line implies normally distributed data.

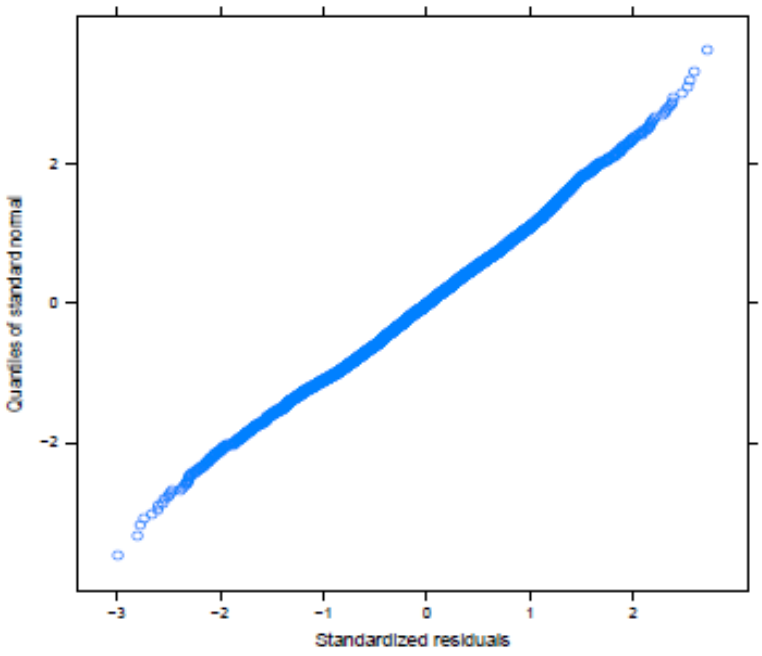


Table 19. Coefficients and significance of the terms used in linear mixed model (analysis of the square-root of the serological response).

Element	Value	p-value	Significant?
Intercept	6.85	<0.0001	Y
City	0.04	0.6885	N
prePost	0.85	<0.0001	Y
p2vp1	0.36	<0.0001	Y
p4vp3	0.55	<0.0001	Y
CityByPrePost	-0.30	<0.0001	Y
CityByp2vp1	0.05	0.54	N
CityByp4vp3	0.09	0.32	N

Most of the single variable effects in Table 19 (*prePost*, *p2vp1* and *p4vp3*) are significant. However, only one of the interaction effects (*CityByPrePost*) is statistically significant ($p< 0.0001$). Since *CityByp2vp1* is not significant ($p=0.54$), it suggests there is no significant difference between Dundee and Glasgow participants in how their serological responses change, between collection periods 1 and 2. A similar conclusion applies to *CityByp4vp3*, which compares the difference in serological positivity between periods three and four by city cohort. Since these two interaction contrasts are insignificant at this stage, we can exclude them in the second linear model based on logarithmic transformed PPR data.

The coefficients for the second linear model fitted to the log of the serological responses are shown in Table 20. The components re-analysed in this second linear model are all those that were statistically significant in the first linear model. The reason for refitting the model with a log transformation is that the parameter estimates are easier to

interpret as percentage changes on the PPR scale. All future analyses would use the log transformed PPR.

Table 20. The coefficients and significance of the terms used in the second linear mixed effects model fitted to the log of the PPR.

Element	Value	p-value	Significant	Coefficient exponent	% difference
Intercept	3.50	<0.0001	Y	33.24	-
City	0.05	0.122	N	1.05	10
prePost	0.30	<0.0001	Y	1.35	82
p2vp1	0.19	<0.0001	Y	1.21	46
p4vp3	0.16	<0.0001	Y	1.18	39
CityByPrePost	-0.17	<0.0001	Y	0.84	-29

Further evidence of an upward trend in serological responses over time can be inferred from the following results. We first consider the effects linked with only the collection periods. From the *p2vp1* contrast, the average serological response during collection period two was greater than period one (for the entire study cohort) by 46% (Table 20). Similarly, the average serological response during period four was greater than period three (for the entire study cohort) by 39% (*p4vp3*). The average serological difference between the pre-filtration (p1+p2) and post-filtration (p3+p4) period was 82% for donors in Dundee, Table 20; in view of the interaction term, *CityByPrePost*, this is the average percentage difference at the lower level for *City* i.e. in Dundee.

Looking at the *City* contrast alone, on average, Glasgow participants had a higher serological response (10%) compared with Dundee donors in the pre-filtration period only, because of the *prePost* interaction. However, this was not itself a statistically significant

difference. This difference itself is a result of the interaction contrast, *CityByPrePost* i.e. between city and time period variables. After the effects of *City* and *PrePost* have been considered independently, the *CityByPrePost* analysis assesses the *PrePost* effect in Glasgow. This showed that serological responses in Glasgow alone were reduced by 29%, compared with Dundee after filtration (Table 20). The changes in serological response implied by the contrasts can best be conveyed graphically in figure 16.

Figure 16. Change in serological responses among the study cohorts by city (Glasgow and Dundee) over time.

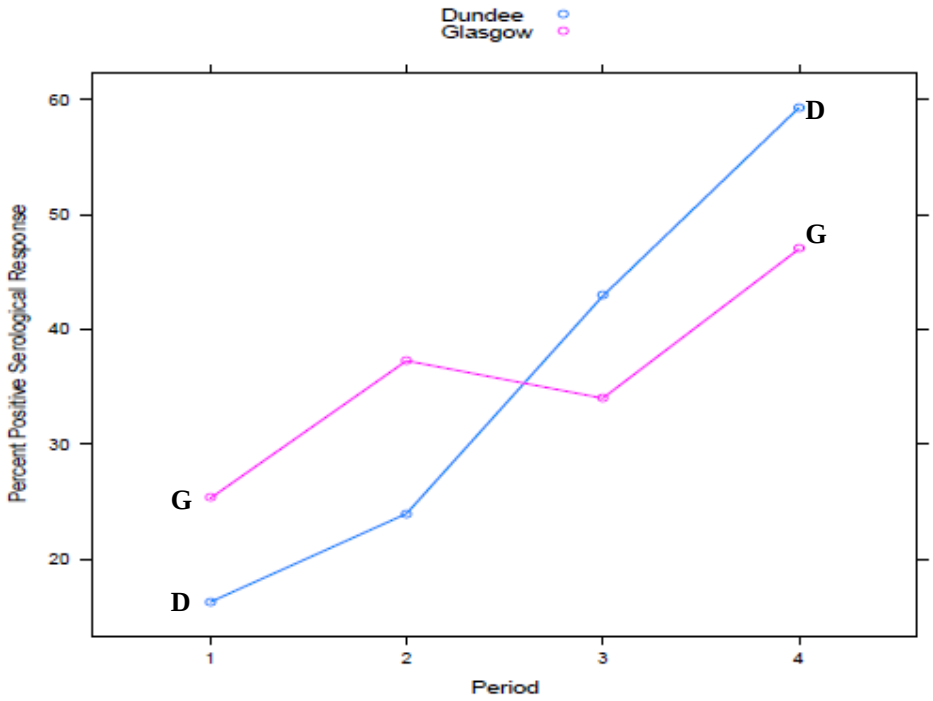


Figure 16 shows the average percent positive serological responses that has been modelled and fitted to Glasgow and Dundee donors. The average is calculated from all donors responding at each time period, in each city. In Glasgow in period 1, the average percent positive serological response among all donors is 26% relative to the positive rabbit control. This is a measure of the serological response within any individual donor and is

not to be interpreted as meaning that 26% of donors in Glasgow have a positive serological response in period 1. The corresponding level for Dundee donors is less than 20% in period 1.

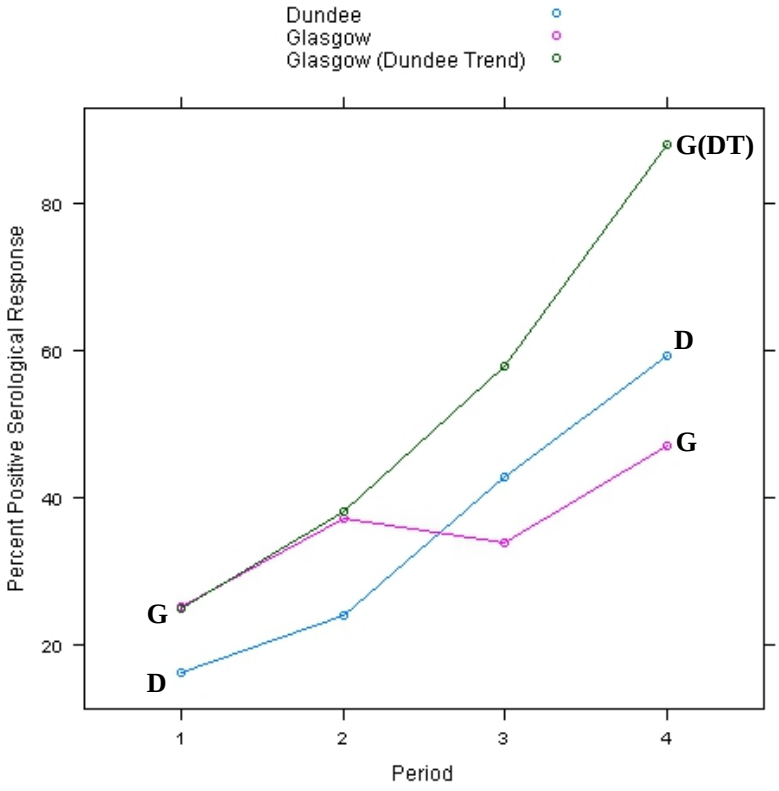
The upward trend of serological responses over time in the Dundee study cohort is clear – there is a consistent increase from one period to the next. In Glasgow there is evidence that the period on period increase is not consistent and the pattern is different from that in Dundee, which is consistently upward (this is what the *CityByPrePost* interaction term in the model above in Table 20 is measuring). In the pre-filtration period (periods 1 and 2) the serological response is higher in Glasgow compared to Dundee but the increase from period 1 to period 2 is similar.

The most interesting change occurs from the pre-filtration period (periods 1 and 2) to the post-filtration period (periods 3 and 4) where there is a step-change in the average serological response in Glasgow with a net reduction compared with Dundee. In both periods 3 and 4, the average serological response is lower in Glasgow compared to Dundee, though the increase from period 3 to period 4 is similar in both cities. This finding further supports the conclusion that the study provides evidence for an association between the introduction of filtration of drinking water from Loch Katrine at Milngavie WTW and the level of the serological response to *Cryptosporidium* in Glasgow donors.

The reduction in the average level of serological response to *Cryptosporidium* could be interpreted as an indication of an overall reduction in exposure to oocysts between periods 2 and 3 among Glasgow study participants. Such a reduction in exposure would be consistent with a reduced risk of encountering oocysts via drinking water from the Loch Katrine supply. This apparent reduction in seropositivity might therefore give an approximation of the amount of seropositivity associated with previous exposure to oocysts via unfiltered Loch Katrine water. However, as can be seen in Figure 15, this effect (if it was associated with the change to water treatment at Milngavie WTW) appears to have been short-lived as the average level of serological response among Glasgow participants

then later increased again, by a similar amount to those donors from Dundee, between time periods 3 and 4.

Figure 17. Change in serological responses among the study cohorts by city (Glasgow and Dundee) over time with predicted trend for Glasgow.



In figure 17, the data for figure 16 plots have the predicted serological response in Glasgow, assuming that the same trend is observed in Dundee (over the 4 periods) and as seen in Glasgow (over periods 1 and 2). These predictions are obtained by fitting a model to the log (PPR) in which there is a term to distinguish the two cities and a linear trend for time. The data from Glasgow in periods 3 and 4 is not used to estimate the parameters of this model. When converted back to the PPR scale in figure 16, the linear trend of the log (PPR) scale becomes a curve on the PPR scale. The predictions for Glasgow in periods 1 and 2 correspond to the observed data and indicate that the model fits the data well and is

therefore robust as a prediction tool. The predictions for Glasgow in periods 3 and 4 represent a 'best estimate' of what would have occurred in Glasgow if filtration had not been introduced, assuming that the trend observed in Dundee carried on in the same way in Glasgow. This is, of necessity, an untestable assumption.

The predicted mean serological response for Glasgow in period 3, was 58% compared to an observed mean serological response of 42%; in period 4 it was 88% compared to an observed mean serological response of 60%. In period 3, the observed mean was 72% of the predicted value, and in period 4 it was 68% of the predicted value, yielding an estimated reduction in serological response associated with the introduction of the filtration plant of 28% in period 3, and 32% in period 4 relative to the predicted mean, assuming that it is valid to extrapolate the Dundee trends post-filtration, to Glasgow.

3.13 Discussion of mixed models and serological responses

Having independently appraised donor serological responses, demographics and behaviour, the use of mixed models (linear and logistic regression) allows a more complex analysis of the interaction of such factors. The logistic model required serological responses to be converted to a positive/negative response before the linear modelling was applied. We appraised a number of 'cut-off' points and were interested in the proportion of donors who were serologically positive for *Cryptosporidium* and how this varied using such 'cut-offs'. Other *Cryptosporidium* seroprevalence studies have used a 20% 'cut-off' threshold to designate 'positivity' relative to their positive controls (but often using sera from clinical cases rather than immunised rabbits) [14, 15, 18]. In such studies, the seroprevalence of positive responses to *Cryptosporidium* ranged from 70-98% of the study populations. However, such comparisons were made with a positive control derived from clinical cases of cryptosporidiosis i.e. antibodies from a human case who had cryptosporidiosis. Given this technical difference, we performed extensive sensitivity

testing to establish an appropriate, scientifically justifiable 'cut-off'. Using a 30% 'cut-off', our entire study cohort had an overall seroprevalence of antibody to *Cryptosporidium* of 64%. If the 'cut-off' was reduced to 20%, then 75% of the entire cohort had a 'positive' serological response to *Cryptosporidium*. This is comparable to other studies.

Both the linear and logistic mixed models provided strong evidence that the proportion of study participants with serological responses in both Glasgow and Dundee donors were increasing over the study period (approximately 3.5 years). However, we noted the decrease in serological responses in Glasgow participants between time periods two and three, when the new filtration system for the Loch Katrine supply was introduced at Milngavie WTW. Therefore, we conclude from these models that there is good evidence, which strongly suggests an association between a change in the seropositivity to *Cryptosporidium* oocysts among participants living in an area supplied by Loch Katrine and the introduction of enhanced filtration treatment to that supply. The introduction of enhanced physical treatment resulted in a 29% reduction (Table 20) in average serological responses among Loch Katrine supply (Glasgow) participants compared with their Clatto (Dundee) counterparts. However, of note was the fact that this effect was short-lived with the average serological responses among Glasgow participants once again increasing in the 4th collection period.

The net effect of the step-change in seropositivity rates among the Glasgow participants was a re-setting of the baseline level of seropositivity to a lower level than in period 1 and 2. This gives a possible indication of the proportion of background seropositivity (and possibly immunity) to *Cryptosporidium* associated with low-level continuous exposure to oocysts via unfiltered drinking water. It suggests that approximately 29% of all exposures to *Cryptosporidium* among the Glasgow participants was attributable to unfiltered drinking water.

These interpretations are however based only on the preceeding analysis, which does not allow for the possibility of interactions btween PPR and changes in the risk factor

profiles of the participants. It should also be considered that there may be other changes, which might affect the risk of exposure to *Cryptosporidium* oocysts in the environment. These will now be considered in more detail.

3.14 Risk factors and their interacting effect on the serological response (PPR)

To consider how risk factors for exposure to *Cryptosporidium* may affect serological responses, we analysed the questionnaire responses from participants in period one. 1050 responses were considered after excluding any second response from participants who donated more than one blood sample in the first period and also any questionnaire with no matching recorded serological sample. Some participants did not answer all questions, leading to a small reduction in total, useable responses. The structure of a question (which was the explanatory variable) determined which statistical technique was used e.g. a binary (yes/no) response was analysed using a t-test. Other questions had an ordinal response e.g. frequently, regularly, occasionally etc. In such instances, a linear model with an ordered factor was used, to test for serological differences. Donors were also assessed for the effect of age, a continuous variable; using a linear model. As repeated tests were carried out on the same variable (the serological response), the level used to assess statistical signficance was corrected using Bonferroni's correction. As there were 31 questions asked, this indicated that a corrected significance level of $p < 0.0016$ should be used as opposed to the standard 0.05. Based on the corrected significance test, there were no statistically significant results in the t-tests carried out on any of the binary response questions.

Although not statistically significant, four responses were of note (some perhaps predictable, other not). Participants with pets had (perhaps surprisingly) a lower serological response to *Cryptosporidium* than those without pets ($p = 0.034$); higher serological

responses were noted for those who reported swimming in the UK (p =0.091); higher serological responses were associated with consuming water from a private water source (p =0.087) (as might have been expected); and lower serological responses were associated with reported drinking of bottled water (p =0.051). Hence, there was no evidence of a signficant relationship at the multi-variate level between serological response and purported risk factors such as swimming, consumption of water from a private water supply etc.

The finding that participants who owned pets and had a reduced serological response to *Cryptosporidium* compared with those who didn't have pets, is of particular interest. The majority of domestic pets are cats and dogs which do not routinely share zoonotic infections with healthy immunocompetent people. Furthermore, although dogs can host *Cryptosporidium canis* and cats are able to host *Cryptosporidium felis*, carriage rates do tend to be low (5-9%) [19]. However, there is some evidence to suggest that pet owners may be exposed to relatively high doses of allergens through dander, which can promote immune tolerance and a resultant reduction of antibody subtypes, which are considered to be inflammation-driven [20, 21]. It is conceivable that the IgG antibody to *Cryptosporidium* 27-kDa antibody may be one of these downregulated antibodies, as the host immune response to *Cryptosporidium* infection tends to be driven by the inflammatory arm of the immune response, often involving IgG sub-type antibodies.

Age as a risk factor

To assess the importance of age on participant serological response, a linear model was fitted to the square root of the serological response (in order to normalise the distribution). In this model, age was associated with a statistically significant difference in the serological response (p <0.0001) i.e. for each year of age, the average serological response increased by approximately 0.35%.

Analysis of variance (ANOVA) was also used to assess if there were statistical differences between groups in the ordinal response questions; e.g. question 10b (Appendix 2) asked "How much bottled water do you consume?" The first contrast checks for a linear relationship between the amount of bottled water consumed and the serological PPR. The next contrast assesses a quadratic relationship between increasing levels of bottled water consumption and the level of serological response. Responses for questions 9c, 9d and 10b of 'none' were excluded to facilitate a clearer investigation of the type of relationship, if any, between amount of water consumed and serological response.

The main results from the application of ANOVA to these three questions showed no significant differences in serological response linked with responses to these ordinal questions (including how much bottled water was drunk) (both p-values > 0.05 and Bonferroni-corrected 0.0013) for the first collection period. However, in view of the earlier analysis suggesting a non-signficant association with bottled water consumption, we considered the effect of bottled water on serological response in more detail, using a linear model. The coefficients for this model are shown in Table 21.

Table 21. Linear fit model to square root of serological response for Q. 10b.

Question 10b. Amount of bottled water consumed	**Value**	**p-value**	**Significance**
Intercept	5.58	-	-
Bottled water frequency (linear)	0.52	0.03	No
Bottled water frequency (quadratic)	-0.24	0.30	No

This model was fitted to those donors who reported drinking at least some bottled water ('None' category was excluded; other options included 1-2 glasses and 2 or more glasses). Although the linear trend contrast (Bottled_water_freq.L) had a p-value of 0.028, it was not statistically significant using the corrected significance value ($p < 0.0013$).

The risk factor components of 'age' and 'consumption of bottled water', were assessed to determine if these two independent factors interacted to modify the donor serological responses (Table 22).

Table 22. Linear fit model assessing interaction of age and bottled water consumption.

Age and Q10.a (Drink bottled water?)	Value	p-value	Significance
Intercept	4.25	-	-
Age	0.04	0.002	No
Bottled_waterYes	-0.27	0.73	No
Age:Bottled_waterYes	0.0006	0.97	No

The coefficient of the interaction term between age and drinking bottled water (AGE:Q10a.Bottled_waterYes) was several orders of magnitude smaller than the other coefficients, suggesting there is little difference in the influence of age as a factor between those who consume bottled water and those who do not. This is reinforced by the interaction term not being significant.

In summary, there is little statistically significant evidence for risk factors for exposure to *Cryptosporidium* oocysts affecting serological responses among donors (in either cohort) in the first collection period. The only donor factor which is significant is age. As people age, they are progressively more likely to encounter micro-organisms such as *Cryptosporidium* from a variety of environmental and domestic sources, which in turn results in the probable augmentation of the immune response and increases the chances of detecting antibodies in a random blood sample.

We have also highlighted some known risk factors for exposure to oocysts which, while not statistically significant in this collection period, may yet be significant in the mixed effects modelling assessing factors across all four collection periods.

3.15 Risk factors, serological response; linear mixed model

Further analysis was carried out on the combination of the serological responses detected and the exposure factor questionnaire responses. We ascertained if different risk factors were associated with variation in serological responses from previous analyses and considered only those risk factors that were shown to have an impact (though not necessarily significant) on serological responses; i.e. age of donors; pet ownership; swimming within the UK; bottled water consumption and the amount of bottled water consumed.

The linear effects mixed model corroborates the previous finding that as age increases, so does the serological response. However, this does not of itself negate the previous conclusion regarding association between serological response and exposure to filtered vs unfiltered water. Pet ownership and swimming within the UK did not affect the previous finding that filtration altered serological responses.

To further explore drinking water habits, we present the drinking water characteristics of participants in Table 23.

Table 23. Drinking water preferences of blood donors. NA denotes questionnaires with no response to whether tap water or bottled water was consumed.

Water Source	None	Tap	Bottled	Both	NA	Total
Frequencies	128	1121	328	2014	115	3706
%	3	30	9	54	3	100

Most participants reported drinking both unboiled tap water and bottled water (54%, 2014/3706). Fewer donors consumed only unboiled tap water (30%, 1121/3706) and even fewer drank only bottled water (9%, 328/3706).

Using the interactions model, we further assessed how drinking water habits and the concomitant donor serological responses were affected by the introduction of the filtration system. In the interaction model, we assessed for the effect of filtration implementation in Glasgow and how this combined with the sources of water consumed, by considering how water source interacted with the *CitybyPrePost* contrast. The coefficients of all of these potential interactions were negative, providing evidence of the new filtration at Milngavie WTW reducing serological responses among the three groups of water consumers in the Glasgow cohort (unboiled alone, bottled water alone and the consumers who drank both types of water). However, donors who consumed any bottled water (either alone or in combination with drinking unboiled water) exhibited the largest reduction in their serological responses. It therefore appeared that participants who consumed any bottled water were relatively more affected in their antibody response reduction that those who drank only tap water. However, there was no evidence for a direct statistically significant relationship between the reported quantity of bottled water consumed and a reduced serological response.

The finding of bottled water consumption and its effect on serological responses appears to be inconsistent with the study hypothesis. It would be reasonable to postulate that a greater reduction in serological response to *Cryptosporidium* might occur in those donors who only consumed unboiled tap water, after filtration was introduced to Milngavie WTW. Moreover, if drinking unfiltered water contributed to *Cryptosporidium* seropositivity in donors and that immunological stimulus of oocysts was then removed by filtration, the antibody response to *Cryptosporidium* would be expected to decline over time and the effect might be expected to be more exaggerated in participants who reported drinking only unboiled drinking water from Loch Katrine. However, the fact that there was no direct relationship between the reported quantity of bottled water consumed and a decrease in seropositivity post-filtration, in the Glasgow donor cohort, suggests the

relationship between seropositivity to *Cryptosporidium*, drinking water and other 'risk factors' may be more complicated than previously considered.

3.16 Linear mixed modelling of the 'original' cohort

In order to improve the rigour of previous analyses and remove the possibility of interactions between factors associated with the loss of some original participants and their replacement by new recruits, further analysis was performed. The most 'stable' participant group were those who were recruited as part of the 'original' cohorts (Dundee and Glasgow) and who continued to donate samples throughout the four periods of the study. We repeated the previous analysis of linear mixed models on serological data from this 'core' cohort of study participants who donated at least once in each of the four collection periods. Of a total of 1437 individual participants who donated at least once during the study, only 269 donated at least once in each collection period (approximately 20% of the 'original' total study cohort).

The mixed modelling results were similar to those of the full combined cohort in that there was an increase in mean serological responses over time, in both donor cohorts (Glasgow and Dundee). Furthermore, this model also confirms that with the introduction of filtration at the Milngavie WTW, the serological responses of Glasgow donors decreased in comparison with Dundee donors (figure 18). Both donor cohorts however, showed a further increase in mean seropositivity from period 3 to period 4 (as before using earlier examples) (figure 18).

Figure 18. Change in serological response over time and by city by 'core' cohort.

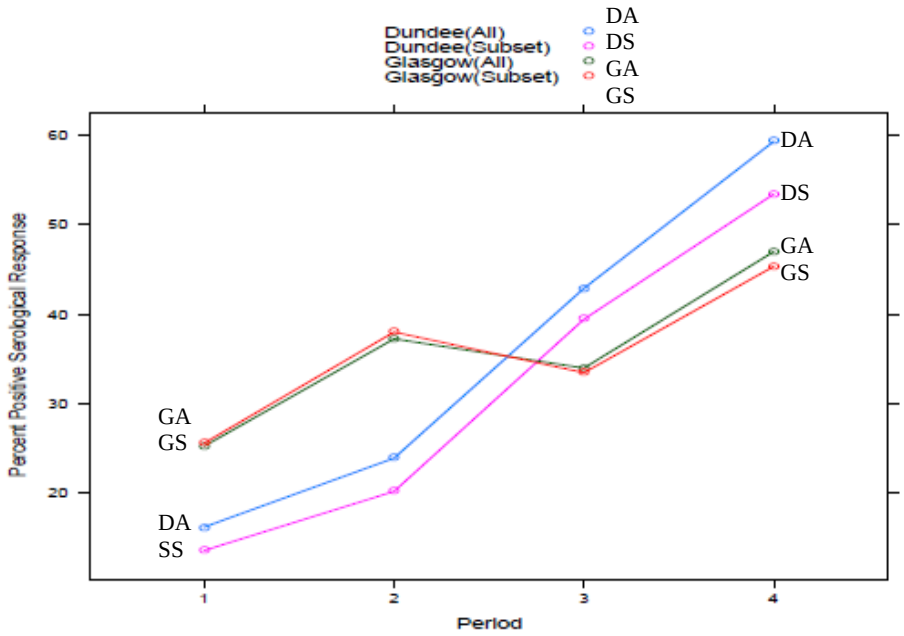


Figure 18 compares the seropositivity (PPR relative to the positive rabbit control) in donors from the Glasgow and Dundee cohorts in total over time with the equivalent data for the 'core' cohort, the subset of donors who donated throughout the study (period 1 to 4 inclusive). This shows that donors in the 'core' cohort recruited at the start of the study, had a slightly lower PPR among Dundee participants compared to those in the 'original' Dundee cohort (period 1 only). In contrast, the mean serological response (PPR) between the Glasgow 'core' cohort and the entire Glasgow cohort was no different. While these data also corroborate the finding that filtration contributes to a reduction (albeit temporarily) in serological responses to *Cryptosporidium* in Glasgow, it reassuringly confirms the addition of new Glasgow donors to the study had no significant bearing or bias on the results overall and that the analyses of the entire Glasgow and Dundee cohort including new replacement donors, is valid.

4 Discussion

Studies of *Cryptosporidium* infection across the world report a wide range of detected serological responses to *Cryptosporidium* oocysts. Paired city surveys have found that consumers of surface water generally have a higher serological response than consumers of groundwater [13, 14]. Such evidence of elevated serological responses appears to confer protective immunity to consumers of surface water. In waterborne outbreaks of cryptosporidiosis, permanent inhabitants routinely exposed to waterborne oocysts appeared to be relatively more immune to cryptosporidiosis whereas visitors experienced illness [3, 22]. Levels of *Cryptosporidium* oocysts in most drinking waters are considered to be low but exposures may also be frequent; chronic low level drinking-water exposures may therefore offer a degree of protection against infection with cryptosporidiosis [15].

Water industry and public health responses in Scotland and the UK generally, to *Cryptosporidium* contamination of drinking water have generally focused on establishing effective multiple barrier water treatment systems, in an effort to eliminate this waterborne pathogen from drinking water supplies. Previous evidence has suggested an association between consumption of unfiltered water from Loch Lomond and the incidence of cryptosporidiosis [10]. In 1999, the addition of rapid gravity filtration and coagulation to the Blairlinnans WTW (Loch Lomond system) was associated with a substantial reduction in the number of confirmed cases of cryptosporidiosis in Central Scotland thereafter.

In 2000, an outbreak of cryptosporidiosis occurred among Glasgow environs residents who received drinking water from Loch Katrine [23]. In order to eliminate *Cryptosporidium* from Loch Katrine-sourced water, enhanced water treatment was introduced to the Loch Katrine supply system at Milngavie WTW in September 2007. One of the aims of this study was to determine if there was any appreciable change in the

background levels of seropositivity to *Cryptosporidium* in Glasgow associated with the introduction of filtration to the Loch Katrine supply.

Using a number of statistical models, we detected a reduction in the seroprevalence to *Cryptosporidium* among blood donors in the Loch Katrine supply area, after the introduction of enhanced filtration treatment. There was no equivalent change among blood donors receiving drinking water from Clatto reservoir, Dundee. Furthermore, the collated evidence suggests that this effect was mainly, if not solely attributable to the introduction of filtration at Milngavie WTW. Further investigation of oocyst rates in raw and final water samples did not make it possible to ascertain whether the change in seropositivity was solely due to the efficacy of the filtration system or whether up-stream practices/change in sampling location may also have contributed to this effect. Although the effect on *Cryptosporidium*-specific antibody appeared to be transient, continued surveillance of clinical cases of *Cryptosporidium* spp. should permit an improved understanding of the long-term consequences of enhanced filtration treatment on Loch Katrine-sourced water, in the Glasgow area population.

Recent evidence suggests that some cryptosporidiosis outbreaks may now be partially due to the success of water suppliers in reducing oocysts to very low concentrations [18]. By reducing low-level waterborne exposures that might otherwise confer protective immunity, improved drinking water treatment in recent years may be a factor in changing the epidemiology of cryptosporidiosis i.e. a common infection with little apparent illness in the past has become a more severe form of disease, which more often necessitates medical attention (McAnulty *et al.*, 2000). This has prompted a rise in GP consultations, stool submissions and a concomitant increase in cases of cryptosporidiosis. However, the need to eliminate sporadic high-dose exposures remains an important public health activity and this is best executed by improving the microbiological quality of drinking water.

It was interesting to observe that Glasgow and Dundee blood donors did not differ significantly in terms of demographics or putative risk factors. Although Glasgow donors consumed more bottled water (and drank less unboiled tap water) than their Dundee counterparts, this did not have an effect on the serological responses. This may be due to the fact that there are a number of other potential transmission pathways for *Cryptosporidium* and/or its antigens, which may stimulate the donor antibody response. For example, in this study, donors who inadvertently consumed water while swimming in a chlorinated, swimming pool or consumed tap water from a private supply were more likely (though not statistically significantly) to have higher serological responses to *Cryptosporidium* than those who did not. The protective effects of such antibody levels are unknown since clinical cases of cryptosporidiosis have nonetheless been associated with swimming pools and private water supplies [24, 25]. More frequent water use may increase the overall level of oocyst exposure but it may also confer some resistance to infection on a per-event basis [30]. As the World Health Organisation noted in its recreational water guidelines, the risk of "infection or disease depends upon the specific pathogen, the form in which it is encountered, the conditions of exposure and the host's susceptibility and immune status" [31].

In this survey, 64-75% of the cohort (depending on chosen positivity cut-off) of the cohort had a strong serological response to the 27-kDa oocyst protein. These findings are consistent with other populations served by surface-water sources [14,18]. There are several aspects to this study which advance previous findings; the study involved a large number of donors followed over a considerable period of time; there was a defined intervention in one population but not the control population; and we sought data on demographics and biologically plausible exposure risk factors for each blood sample donation. Furthermore, this was a prospective cohort study; such studies are considered to be one of the most robust epidemiologic methodologies.

Other serological studies have focused on IgG serological responses to both the 15/17-kDa complex and the 27-kDa protein [13-15, 17, 18]. After exposure to *Cryptosporidium*, a serological response to both of these antigen groups usually peaks 4-6 weeks later [26]. The 15/17-kDa marker declines to baseline levels in 4-6 months while the 27-kDa marker remains elevated for at least 6-12 months. However, we recognise that a potential weakness of the study was that we were not able to investigate serological responses to the 15/17-kDa complex. However, we do not believe this undermines the main findings of the study as the 27-kDa response is considered a reliable marker for exposure to *Cryptosporidium* and we were primarily interested in evidence of oocyst exposures over lengthy periods, not in detecting evidence of recent, acute infection

This was surprising in that part of the original hypothesis (and a reason for including a second water supply population who had long-term experience of consuming only filtrered drinking water) was that the background level of *Cryptosporidium* antibodies might be relatively stable over time in a population. The study has demonstrated unequivocally that in the donor populations sampled in both areas, this was not the case. For both Glasgow and Dundee donors, the level of seroprevalence increased over time. For every year of life, this resulted in a 0.35% increase in serological PPR to *Cryptosporidium*.

The current study suggests that sources other than drinking water may frequently transmit *Cryptosporidium* and be a source of oocyst exposure. In fact, food, zoonotic and other modes of parasite transmission may be at least as important as drinking water and may be more likely to transmit higher dose oocyst exposures. The observation that age correlates with an increase in serological response to *Cryptosporidium* therefore should not be too surprising as this has also been observed for other gastro-intestinal pathogens as well as other infections [28, 34]. This may also explain why children are much more susceptible to gastro-intestinal infection (including cryptosporidiosis) than adults, as adults are much more likely to have quantitatively higher serum/mucosal antibody levels induced by a number of pathogens and exposures during the life experience, than children [29].

However, unlike other gastro-intestinal pathogens, *Cryptosporidium* is an organism which is resistant to all but the harshest of environmental conditions. The oocyst is likely to remain infectious in the environment for much longer than other gastro-intestinal pathogens, resulting in an increased risk of exposure to this particular organism.

Taken at a whole population level, the background rate of seropositivity in likely to be stable. As older people with relatively higher antibody levels die, they are replaced by newborns with no antibodies. There will therefore be an 'equilibrium level' for the total population. However, the ability of any given population to resist a sudden increase in oocyst exposure and hence the probability of an 'outbreak' of cryptosporidiosis must be related to this 'equilibrium' level of oocyst antibody response. The study was not designed to assess this.

Persons with previous or on-going exposure to *Cryptosporidium* develop partial (protective) immunity and contamination of community water sources may not necessarily manifest as detectable outbreaks among residents. Dramatic rates of community-wide disease may not be the inevitable result of on-going chronic low level water supply contamination [3]. The complete removal of pathogens from drinking water may decrease the risk of waterborne enteric illnesses associated with them but may also increase the risks from other exposures to the same pathogens. It would be apposite to perform a retrospective study of all sporadic microbiologically-confirmed cases receiving Loch Katrine water before and after the introduction of the enhanced filtration at Milngavie WTW to investigate further the long-term clinical impact on the local population of the introduced filtration system.

The study was specifically designed to detect evidence of any association between the levels of antibody to specific *Cryptosporidium* antigens and changes to the treatment of drinking water sourced from Loch Katrine, before and after a new filtration system was introduced in September 2007. The study has convincingly demonstrated that there was robust evidence of just such an association; that the impact of improving water treatment at

Milngavie WTW was to reduce the level of background antibody levels in local consumers but that this effect was only temporary. The other finding of note, which wasn't predicted, was that the level of antibodies to *Cryptosporidium* in a defined cohort increases with time, as that cohort grows older. Age itself is therefore strongly associated with increasing levels of *Cryptosporidium* antibody to the 27-kDa outer wall protein.

The underlying hypothesis behind the study was that by consistently improving the quality of drinking water treatment in Scotland, the population may inadvertently be rendered more vulnerable to sporadic infection with cryptosporidiosis from other sources or worse, exposing the population to larger outbreaks of cryptosporidiosis if water treatment systems failed and allowed oocysts to enter the drinking water supply. The 'step-change' reduction in the antibody levels in the Glasgow donor cohort was transient and the underlying upward trend due to ageing, suggests that populations drinking oocyst-free drinking water are not in fact likely to be significantly more at risk of *Cryptosporidium* infection than if drinking unfiltered tap water. Any increased vulnerability is likely to associated with reduced antibodies with advancing age and the medications, which may reduce CD4 T cell counts e.g. chemotherapy/steroid treatment for rheumatoid arthritis etc.

5 Conclusions

The hypothesis of an association between the introduction of enhanced physical treatment (filtration) of the Loch Katrine water supply in September 2007 and a reduction in background levels of antibody in the local consumers, to *Cryptosporidium*, was conclusively proven. The magnitude of this effect was significant, amounting to the equivalent of approximately a 29% reduction in antibody positivity in Glasgow donors, compared to Dundee donors, over the period immediately following the introduction of filtration treatment in September 2007. This change was however a one-off transient event, not a sustained cumulative reduction. After the immediate impact, the overwhelming determinant of antibody levels was correlated with the impact of ageing in the donor cohorts. Further conclusions are:

- Glasgow and Dundee donors did not differ significantly in terms of demographics or putative risk factors.

- In this survey, between 64-75% of the cohort (depending on chosen cut-off) had a detectable serological response to the 27-kDa oocyst protein. These findings are consistent with other populations served by surface-water sources.

- The current study suggests that sources other than drinking water may frequently transmit *Cryptosporidium* and be responsible for background levels of antibody response to oocyst exposure.

- We detected a change in the seroprevalence to *Cryptosporidium* in the Glasgow blood donor population after the introduction of enhanced physical treatment. This effect was not observed for the Dundee population. Furthermore, the collated

evidence suggests that this effect was solely attributable to the introduction of filtration at Milngavie WTW.

- Any impact on population susceptibility to cryptosporidiosis associated with introducing enhanced water treatment is likely to be minor, transient and dominated by the effects of ageing and continuous low level exposure from other natural sources.

6 Future research options

The present study used a self-selected population of blood donors who may not necessarily be completely representative of the population as a whole. Therefore, in order to fully understand the long-term consequences (clinical and population level) of abrogating environmental risk factors such as *Cryptosporidium* through improved drinking water treatment such as filtration, it will be important to continue surveillance of clinical infection with *Cryptosporidium* spp. As drinking unfiltered water from Loch Lomond has historically been associated with confirmed cases of cryptosporidiosis [10], we hypothesise that consumption of unfiltered water from Loch Katrine may have been associated with the incidence of clinical cryptosporidiosis. To this end, it is our intention to perform a retrospective cohort study of microbiologically-confirmed cases of cryptosporidiosis among residents supplied from Loch Katrine. The period of study (2004 to 2010) will be chosen to enable the detection of any relevant epidemiologic trends in sporadic cases and exposure risk factors. The implementation of the filter at Milngavie in 2007 will be considered during the analysis to determine evidence of any impact on disease incidence before and after the event.

7 Lessons learned

- This was a complex and ambitious study design. The use of a separate control population, as well as a before/after intervention study design in the Glasgow cohort was complicated. The level and sophistication of the statistical analysis required to investigate , analyse and interpret the data was considerable and more than was originally anticipated. Access to very high level post-graduate statistical input proved critical to the successful completion of the study. In any future such study, the earlier involvement of statistical support will be crucial in the design stage to maximise the power of the study and increases the chances of success.
- Selection of an appropriate control sample to compare donor serological responses was a contentious issue and provoked much discussion. Although a control rabbit antibody had been used in other studies, the ideal control sample would have been the pooled sera from clinical cases of human cryptosporidiosis. However, it should be noted that the methodological principles used to calibrate 'positivity' would have been the same. Obtaining sera from clinical cases of cryptosporidiosis is difficult, often requiring ethical approval even if there is a suitable serum bank, which was not the situation in this study. In this magnitude of study, it would also have been extremely difficult to procure the volumes of human sera required.
- Prior to the study, consideration was given to statistical power needed to detect a significant effect in antibody responses if one was evident. In consultation with SNBTS, a donor 'return rate' of approximately 85% (of the original recruited cohort) was suggested and was used for the study. In the course of the study, we found this to be nearer 30%; there was therefore a much higher attrition rate of donors than anticipated and significant recruitment of new donors had to be performed to maintain the study numbers. While addition of new donors to the

study did not affect the demographics, the serology results or risk factor exposures it may have been more prudent to have recruited more donors from the start of the study to allow for the drop-out rate.

Bibliography

1. Hunter PR *et al.* Health sequelae of human cryptosporidiosis in immunocompetent patients. Clin Infect Dis 2004; 39: 504-510.

2. Hunter PR and G Nichols. Epidemiology and clinical features of Cryptosporidium infection in immuno-compromised patients. Clin Microbiol Rev 2002; 15: 145–154.

3. McAnulty JM, Keene WE, Leland DE, Hoesly F, Hinds B, Stevens G *et al*. Contaminated drinking water in one town manifesting as an outbreak of cryptosporidiosis in another. Epidemiol Infect 2000; 125: 79-86.

4. Goh S, Reacher M, Casemore DP, Verlander NO, Charlett A, Chalmers RM *et al.* Sporadic cryptosporidiosis, North Cumbria, England, 1996-2000. Emerg Infect Dis 2004; 10: 1007-1015.

5. Goh S, Reacher M, Casemore DP, Verlander NO, Charlett A, Chalmers RM *et al.* Sporadic cryptosporidiosis decline after membrane filtration of public water supplies, England, 1996-2002. Emerg Infect Dis 2005; 11: 251-259.

6. Smith HV, Robertson, LJ and Ongerth JE. Cryptosporidiosis and giardiasis: the impact of waterborne transmission. Journal of Water Supply: Research and Technology – Aqua 1995; 44(6): 258-274**.**

7. Smith HV, Patterson WJ, Hardie R, Greene LA, Benton C, Tulloch W et al. An outbreak of waterborne cryptosporidiosis caused by post-treatment contamination. Epidemiol Infect 1989; 103 : 703-715.

8. MacKenzie WR, Schell WL, Blair KA, Addiss DG, Peterson DE, Hoxie NJ et al. A massive outbreak in Milwaukee of *Cryptosporidium* infection transmitted through the public water supply. N Engl J Med 1994; 331: 161-167

9. Smith HV & Grimason AM. *Giardia* and *Cryptosporidium* in water and wastewater. Chapter 39, In Handbook of Water and Wastewater Microbiology, 2003.

10. Pollock KGJ, Young D, Smith HV & Ramsay CN. Cryptosporidiosis and filtration of water from Loch Lomond, Scotland. Emerg Infect Dis 2008; 14: 115-120.

11. Juranek DD. Cryptosporidiosis: Sources of infection and guidelines for prevention. Available at http://www.cdc.gov/crypto/prevention.html, accessed 9 June 2009

12. Ungar BL, Soave R, Fayer R & Nash TE. Enzyme immunoassay detection of immunoglobulin M and G antibodies to *Cryptosporidium* in immunocompetent and immunocompromised persons. J Infect Dis 1986; 153: 570-576.

13. Frost FJ, Muller T, Craun FG, Calderon RL & Roefer PA. Paired city *Cryptosporidium* serosurvey in the southwest USA. Epidemiol Infect 2001; 126: 301-307.

14. Frost FJ, Kunde TR, Muller TB *et al*. Serological responses to *Cryptosporidium* antigens among users of surface- vs- ground-water sources. Epidemiol Infect 2003; 131: 1131-1138.

15. Frost FJ, Roberts M, Kunde TR, Craun G, Tollestrup K, Harter L *et al*. How clean must our drinking water be: the importance of protective immunity. J Infect Dis 2005; 191: 809-814.

16. Chappell CL, Okhuysen PC, Sterling CR *et al.* Infectivity of *Cryptosporidium parvum* in healthy adults with pre-existing anti-*C. parvum* serum immunoglobulin G. Am J Trop Med Hyg 1999; 60: 157-64.

17. Elwin K, Chalmers RM, Hadfield SJ, Hughes S, Hesketh LM, Rothburn MM *et al.* Serological responses to Cryptosporidium in human populations living in areas reporting high and low incidences of symptomatic cryptosporidiosis. Clin Infect Dis 2007; 13: 1179-1185.

18. Kozisek F, Craun GF, Cerovska L, Pumann P, Frost F, Muller T. Serological responses to Cryptosporidium-specific antigens in Czech populations with different water sources. Epidemiol Infect 2008; 136: 279-286.

Formatted: English (U.K.)

19. Overgaauw PA, van Zutphen L, Hoek D, Yaya FO, Roelfsema J, Pinelli E *et al.* Zoonotic parasites in fecal samples and fur from dogs and cats in The Netherlands. Vet Parasitol 2009; 163: 115-122.

20. Woodfolk JA. High-dose allergen exposure leads to tolerance. Clin Rev Allergy Immunol 2005; 28: 43-58.

21. Platts-Mills TA, Woodfolk JA, Erwin EA, Aalberse R. Mechanisms of tolerance to inhalant allergens: the relevance of a modified Th2 response to allergens from domestic animals. Springer Semin Immunopathol 2004; 25: 271-279.

22. Frost FJ, Muller T, Craun GF, Fraser D, Thompson D, Notenboom R, Calderon RL. Serological analysis of a cryptosporidiosis epidemic. Int J Epidemiol 2000; 29: 376-379.

23. National Health Service for Scotland, Greater Glasgow Outbreak Control Team. Report of an outbreak of cryptosporidiosis in the area supplied by Milngavie Treatment Works–Loch Katrine water. Glasgow: Department of Public Health, Greater Glasgow Health Board; 2001.

24. Boehmer TK, Alden NB, Ghosh TS, Vogt RL. Cryptosporidiosis from a community swimming pool: outbreak investigation and follow-up study. Epidemiol Infect 2009; 137: 1651-1654.

25. Pollock KGJ, Ternent HE, Mellor DJ, Chalmers RM, Smith HV, Ramsay CN, Innocent GT. Spatial and temporal epidemiology of sporadic human cryptosporidiosis in Scotland. Zoonoses Public Health 2009; epub ahead of print.

26. Moss DM, Chappell CL, Okhuysen PC, Dupont HL, Arrowood MJ, Hightower AW, Lammie PJ. The antibody response to 27-, 17- and 15-kda *Cryptosporidium* antigens following experimental infection in humans. J Infect Dis 1998; 178: 827-833.

Formatted: German (Germany)

27. Priest JW, Bern C, Roberts KM, Kwon JP, Lescano AG, Checkley W *et al.* Changes in serum immunoglobulin G levels as a marker for Cryptosporidium sp. infection in Peruvian children. J Clin Microbiol 2005; 43: 5298-5300.

28. Reymond D, Johnson RD, Karmali MA, Petric M, Winkler M, Johnson S *et al.* Neutralizing antibodies to *Escherichia coli* Vero cytotoxin 1 and antibodies to O157 lipopolysaccharide in healthy farm family members and urban residents. J Clin Microbiol 1996; 34: 2053-2057.

29. Frost FJ, Craun GL, Calderon CL, Hubbs SA. So many oocysts, so few outbreaks. J Am Water Works Assoc 1997; 89: 8-9.

30. Schijven J, de Roda Husman AM. A survey of diving behaviour and accidental water ingestion among Dutch occupational and sport divers to assess the risk of infection with waterborne pathogenic microorganisms. Environ Health Perspect 2006; 114: 712-717.

31. World Health Organisation. Guidelines for Safe Recreational Water Environments. Vol. 1. Coastal and Fresh Water. Geneva, Switzerland: World Health Organisation; 2003.

32. Goldstein ST, Juranek DD, Ravenholt O, Hightower AW, Martin DG, Mesnik JL *et al.* Cryptosporidiosis: an outbreak associated with drinking water despite state-of-the-art water treatment. Ann Intern Med 1996; 124: 459-468.

33. Frost FJ, De la Cruz AA, Moss DM, Curry M, Calderon RL. Comparisons of ELISA and Western blot assays for detection of *Cryptosporidium* antibody. Epidemiol Infect 1998; 121: 205-211.

34. Badami KG, McQuilkan-Bickerstaffe S, Wells JE, Parata M. Cytomegalovirus seroprevalence and 'cytomegalovirus-safe' seropositive blood donors. Epidemiol Infect 2009; 137: 1776-1780.

Formatted: French (France)

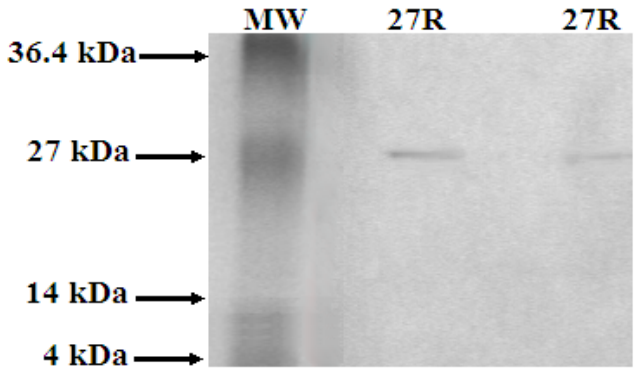
Formatted: German (Germany)

Formatted: French (France)

Appendix 1 – Pilot study of ELISA testing for *Cryptosporidium* antibody detection

During the course of this study, SPDL piloted an ELISA to the 27-kDa protein group. This small-scale study aimed to purify *Cryptosporidium* protein, which had been over-expressed in large quantities from *Escherichia coli*. This recombinant protein was purified using chromatography and the intention was to test donor serological samples in a similar way to the Western blot methodology. One advantage of an ELISA method is that it would allow quicker testing of donor serological responses to *Cryptosporidium* although it is not considered to be as sensitive as Western blot methods [27]. The purified recombinant protein is shown in figure 19.

Figure 19. 12% SDS-PAGE of purified recombinant 27-kDa protein.



MW - Molecular Weight Marker

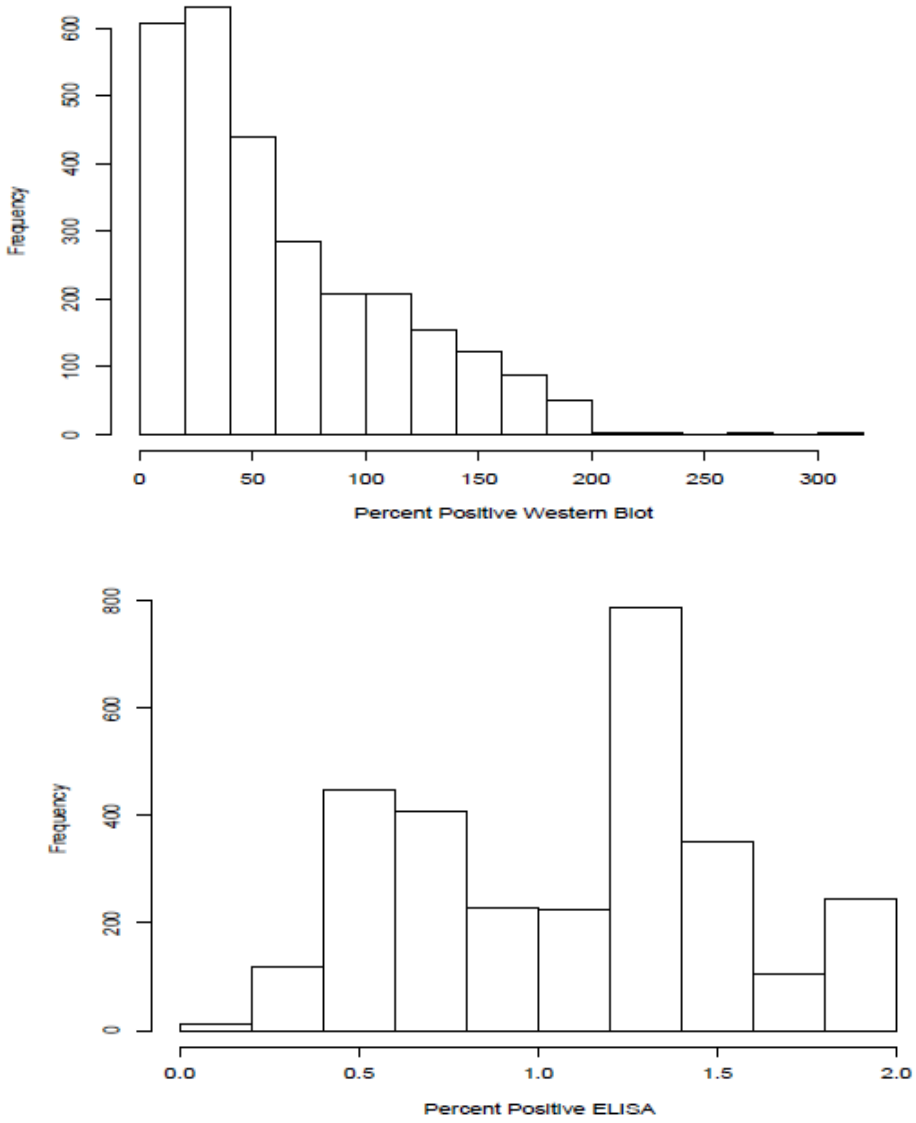
27R - 27 kDa Recombinant Protein

Once the ELISA had been developed and standardised, the two assays were compared using a selection of serum samples which were determined as **positive, negative** and **borderline** by the original Western blot methodology. Preliminary results and statistical analysis showed that there did not appear to be a meaningful correlation between

the two assays. For a more comprehensive comparison, it was decided to test the majority of donor samples from the whole cohort.

Of the full donor cohort, we obtained 2921 (79%) sample results for the ELISA test. Using this subset of all the donations for which we had both test result values, it was possible to compare the methods and see how they relate. Of the 2921 ELISA results, 2747 (75%) were matched to Western blot sample results by unique SNBTS barcode. However, given the quantitative difference in the formats of the results of the two tests, it was not possible to directly compare the two assays. Instead, we considered the percentage positive response (PPR) of each test (donor result value divided by positive control value). The distribution of results for the Western blot and ELISA are displayed in figure 20.

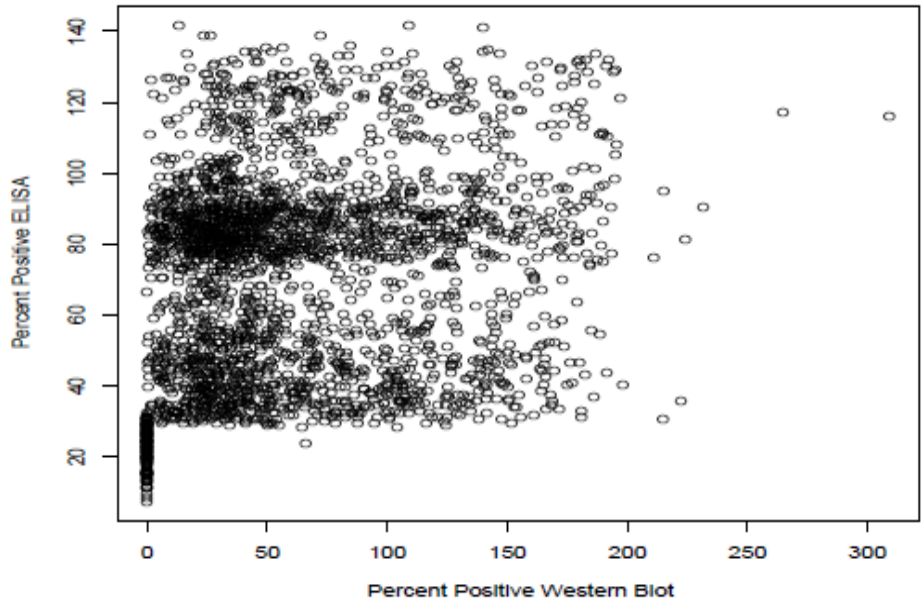
Figure 20. Distribution of percentage positive values for ELISA vs Western blot.



As we have commented previously, the distribution of the Western blot is skewed to responses being less than 50% of the positive control. In contrast, the distribution of the serological responses by ELISA appears to be bi-modal. The uni-modal nature of the Western blot compared with the bi-modal nature of the ELISA suggests that a direct relationship between the two assays is unlikely.

In order to establish whether there was a relationship between the ELISA and Western blot, we used a scatterplot to analyse the data (figure 21).

Figure 21. **Scatterplot of percentage positive results for ELISA vs Western Blot.**



Using scatterplots, it enables us to assess whether the two assays are comparable and have a (diagonally) linear relationship i.e. as one assay detects increases for positivity, so does the other. The graph shows quite clearly that this is not the case. Many low values in the Western blot correspond to high values for donors tested by the ELISA methodology. Even by 'normalising' the data using the square root transformation, the relationship between the data is not significantly improved. Using a statistical technique called Pearson's correlation, the normalised data give a value of 0.32 (95% CI 0.29-0.36) or 32% correlation, $p< 0.001$, which is a weak correlation. A value of 1.00 would represent a perfect linear relationship. Interestingly, this corroborates a similar comparison of the Western blot and ELISA methodology by Frost *et al.* [33], who reported a correlation coefficient of 0.41. This current large Scottish study compared the two assays in order to

fully appraise the best methodology for determining differences (if any) in antibody positivity to *Cryptosporidium*. We determined that the Western blot showed greater statistical separation between donors who were considered to have a positive serological response than those who did not and that this assay is more sensitive than the ELISA. Furthermore, we (and others [33]) believe the Western blot is better at characterising the magnitude of risk of *Cryptosporidium* infection and in evaluating the importance of various risk factors.

Appendix 2 – Epidemiological exposure risk factor questionnaire

RESEARCH INTO MARKERS FOR CRYPTOSPORIDIUM INFECTION

PATIENT QUESTIONNAIRE

All information is strictly confidential

Donor Number (BTS staff only)

Date of birth ☐☐ ☐☐ ☐☐

Age ☐☐

Sex ☐

Full postcode of current residence ☐☐☐☐ ☐☐☐

Length of stay at current residence in years <1 , 1 , 2 , 3 , 4 , 5 , 5+

Are you employed outside the home Yes/No

If yes, please give postcode of main place of work (if known)
or full postal address ☐☐☐☐ ☐☐☐

Length of employment at main place of work <1 , 1 , 2 , 3 , 4 , 5 , 5+

MEDICAL HISTORY

1. Have you ever been diagnosed with *Cryptosporidium* infection?

☐ Yes (if yes, approximately when was this? Month…….Year……)
☐ No

2. Have you had a diarrhoeal illness lasting 4 days or more, within the last 12 months?

☐ Yes (when did this occur? Month……Year……..)
☐ No
☐ Not sure

RISK FACTORS FOR EXPOSURE TO CRYPTOSPORIDIUM

3. a) Do you have children under the age of 5, in your household?

☐ Yes (answer question 3 d)
☐ No (go to question 4)

3. b) Do children in your household attend day care/nursery/playgroup/creche?

- ☐ Yes
- ☐ No

3. c) Have any children in your household been diagnosed with *Cryptosporidium* infection in the last 12 months?

- ☐ Yes (If yes, when was this? Month…….Year……..)
- ☐ No

3. d) Have any children in your household been ill with diarrhoea (lasting 4 days or more) in the last 12 months?

- ☐ Yes (If yes, when was this? Month…….Year……..)
- ☐ No
- ☐ Not sure

4. a) Does your household have pets?

- ☐ Yes (if yes, what type of animal(s)…….)
- ☐ No (go to question 5)

4. b) Have any pets in your household been diagnosed with *Cryptosporidium* infection in the last 12 months?

- ☐ Yes (if yes, when was this? Month……..Year……)
- ☐ No

4. c) Have any of your pets been ill with diarrhoea (lasting 4 days or more) in the last 12 months?

- ☐ Yes (if yes, when was this? Month……..Year……)
- ☐ No
- ☐ Not sure

5. Have you had any physical contact with farm animals, exotic animals or zoo animals in the last 12 months?

- ☐ Yes (if yes, what type of animal(s)……)
- ☐ No

6. a) Have you been swimming in the UK in the last 12 months?

- ☐ Yes (if yes, answer question 6 b to 6 d)
- ☐ No (go to question 7)

6. b) How often do you go swimming normally?

- ☐ Frequently (once a week or more)
- ☐ Regularly (more than once a month but less than once a week)
- ☐ Occasionally (less than once a month)

6. c) When you go swimming, do you swallow any water?

- ☐ Never
- ☐ Sometimes
- ☐ Usually

6. d) In the last 12 months, have you been swimming in any body of water other than a swimming pool?

- ☐ No

☐ Yes (tick all that apply) ☐ River
☐ Loch/Lake
☐ Reservoir
☐ Sea

7. Have you visited anywhere outside the UK in the last 12 months?

☐ Yes (please list countries where visited and when; use extra space at the bottom of page if required……)
☐ No (go to question 9)

8. a) If you have been outside the UK in the last 12 months, did you go swimming while abroad?

☐ Yes (go to question 8 b)
☐ No (go to question 8 c)

8. b) When you went swimming abroad, did you swallow any water?

☐ Yes
☐ No

WATER CONSUMPTION

9. a) Is your water supply private or a public (Scottish Water) supply?

☐ Private
☐ Public

9. b) Do you normally consume un-boiled tap water e.g. drinking, making up diluting squash or in ice cubes ?

☐ Yes (go to question 9 c)
☐ No

9. c) At home, how much un-boiled tap water (or diluting squash etc) do you normally drink, in an average day?

None ☐
Up to one cup/glass ☐
1-2 cups/glasses ☐
More than 2 cups/glasses ☐

9. d) At your normal place of work how much un-boiled tap water do you normally drink, in an average day?

None ☐
Up to one cup ☐
1-2 cups ☐
More than 2 cups ☐

9. e) At home, does your water supply have an in-line domestic water filter fitted?

- ☐ Yes (If yes, how long have you used this? < 1 yr / 1-5 yrs / >5 yrs)
- ☐ No

9. f) At home, do you normally filter your tap water using a jug type filter?

- ☐ Yes (if yes, how long have you used this? < 1 yr / 1-5 yrs / >5 yrs)
- ☐ No

10. a) Do you drink bottled water?

- ☐ Yes (go to question 10 b)
- ☐ No

10. b) How much bottled water do you drink normally, in an average day?

- ☐ None
- ☐ Up to one cup
- ☐ 1-2 cups
- ☐ More than 2 cups

11. Has any other adult or child in your household been diagnosed with Cryptosporidium infection in the last 12 months?

- ☐ **Yes**
- ☐ **No**

Thank you for taking the time to complete this questionnaire.

Please return it to the BTS staff with your signed consent form.