A Case Control Study Investigating Drinking Water & Dairy Products in the Aetiology of Crohn's Disease - A Possible Role for *Mycobacterium avium paratuberculosis* – The CMAW Study

Final Report

Date: 15th July 2005

This report was prepared by:

Ibrahim Abubakar, Senior Lecturer in Health Protection[1]

Deborah J. Myhill, Research Associate in Biomarker Epidemiology[1]

Iain Lake, Lecturer in Environmental Science[2],

Ian Harvey, Professor of Epidemiology[1],

Paul R. Hunter, Professor of Health Protection[1],

[1]School of Medicine, Health Policy & Practice.

[2]School of Environmental Sciences.

University of East Anglia, Norwich, Norfolk. NR4 7TJ.

The full project team includes:

Dr Andrew Hart (chief investigator) and Professor Paul Hunter (chief investigator)

Dr Deborah Myhill, Dr Ibrahim Abubakar, Dr Iain Lake, Professor Ian Harvey, Kate Johnston[3], Professor Jonathan Rhodes[3], Helen Brunskill[4], Dr Richard Robinson[4], Elizabeth Rigney[5], Dr Alan Lobo[5], Imelda Dunn[6] and Dr Christopher Probert[6].

[3]Clinical Trials Department, Royal Liverpool University Hospital, Liverpool, L7 8XP.

[4]Department of Gastroenterology, Leicester General Hospital, University Hospitals of Leicester NHS Trust, Leicester, LE5 4PW.

[5]Department of Gastroenterology, Royal Hallamshire Hospital, Sheffield Teaching Hospitals NHS Trust, Sheffield, S10 2JF.

[6]Department of Gastroenterology, Bristol Royal Infirmary, United Bristol Healthcare NHS Trust, Bristol, BS2 8HW.

Acknowledgement: This work was funded by the Department for Environment, Food and Rural Affairs (DEFRA), UK. The contract was managed by the Drinking Water Inspectorate (DWI), UK.

Disclaimer: Any views expressed in the following report are not necessarily those of DEFRA.

TABLE OF CONTENTS

SUMMARY

This report presents the final results of a case control study investigating the role of drinking water potentially contaminated with *Mycobacterium avium paratuberculosis* (MAP) in the aetiology of Crohn's disease (CD). Patients with CD were recruited from five centres in England and their exposure to proxy measures of MAP contamination was compared with that of controls from a similar geographical area to examine potential risk factors for CD. Exposure to water, milk or dairy products potentially contaminated with MAP will be associated with CD was the primary hypothesis under investigation. Geographical information systems (GIS) software was used to estimate a risk score based on the type of water treatment and source of water supplied to each individual. Odds ratios (OR) for each risk factor were calculated using logistic regression models. Multivariable logistic regression was used to control for the confounding effect of known risk factors. A total of 218 cases and 812 controls were recruited. No significant association was observed between measures of potential contamination of water sources with MAP (surface versus ground water OR 1.08, 95 % CI 0.86 – 1.37), water intake (boiled water OR 0.99, 95 % CI 0.88 – 1.11, or unboiled water OR 1.01, 95 % CI 0.90 – 1.15) or water treatment (sedimentation OR 0.84, 95 % CI 0.31 – 2.28, coagulation OR 0.96, 95 % CI 0.58 – 1.58, flocculation OR 0.73, 95 % CI 0.44 – 1.20 and filtration OR 0.78, 95 % CI 0.51 – 1.18). Neither was there an association with bottled water (OR 0.99, 95 % CI 0.89 – 1.09). Although dairy product intake (OR 0.96, 95 % CI 0.85 – 1.07) was not associated with an effect, consumption of pasteurised milk (OR 0.82, 95 % CI 0.69 – 0.97) was associated with a reduced risk of CD.

In addition to the primary hypotheses tested several other variables remained in the final model. Both smoking (OR 1.31, 95 % CI 1.12 – 1.53) and family history (OR 7.13, 95 % CI 3.37 – 15.08) of CD, two well known risk factors, were significantly associated with an increased risk. In

addition, holidays abroad (OR 0.52, 95 % CI 0.32 – 0.84) and fruit consumption (OR 0.78, 95 % CI 0.67 – 0.92) were negatively associated with risk whilst meat intake (OR 1.40, 95 % CI 1.17 – 1.67) was associated with an increased risk of developing CD. Sub group analysis showed non-stricturing non-penetrating CD to be associated with increased meat intake (OR 1.48, 95 % CI 1.19 – 1.85) while stricturing and penetrating disease were not. CD affecting only the ileocolon was significantly associated with increased meat intake (OR 1.58, 95 % CI 1.26 – 1.98). CD of the upper GI was the only disease site significantly associated with farm holidays (OR 3.55, 95 % CI 1.17 – 10.79) and oral contraceptive use (OR 1.17, 95 % CI 1.01 – 1.36).

This study does not provide any evidence of an increased risk of CD in association with exposure to the two primary hypotheses (drinking water and milk) for transmission pathways of MAP. Consequently this study does not support a role for MAP or drinking water in the aetiology of CD. This conclusion is further supported by the lack of any association with farm holidays and contact with farm animals. Although fruit consumption was not one of the primary hypotheses, the negative association with CD has been noted in previous studies, adding weight to this finding. To our knowledge this is the first study to identify a negative association with travel abroad or a positive association with meat consumption. As these observations were not part of the primary hypothesis care must be exercised in their interpretation until confirmed in subsequent research, as such findings could have arisen by chance. If the association with meat consumption is confirmed, we do not consider that this would indicate an infectious aetiology, but possibly a link with total animal protein consumption, a finding that has been noted previously.

BACKGROUND

Our understanding of the pathogenesis, pathophysiology and clinical management of Crohn's disease (CD) has dramatically evolved over the last decade due to advances in molecular biology and genetics. However, the aetiology is still far from clear. Dalziel in 1913 first observed histopathological and clinical similarities between human chronic granulomatous enteritis (Crohn BB et al. 1932), animal paratuberculosis and intestinal tuberculosis (Dalziel TK 1913). Animal paratuberculosis is caused by *Mycobacterium avium paratuberculosis* (MAP) (Johne HA et al. 1895), while intestinal tuberculosis is caused by *Mycobacterium tuberculosis*. Due to similarities between the diseases, MAP has been suspected to play a role in the aetiology of CD. Published studies using molecular, serological and other microbiological methods have shown some evidence in support of this hypothesis (Bull et al. 2003;Dell'Isola et al. 1994;Elsaghier et al. 1992;Moss et al. 1992;Naser et al. 2004;Naser et al. 2000). However due to inconsistencies in the data, there is currently no consensus as to whether a causal link between MAP and CD exists. Various systematic reviews have been conducted which conclude that a causal link cannot be established based on current evidence, but suggest further epidemiological research is needed (Andersen et al. 1997;Chiodini et al. 1996;Hermon-Taylor et al. 2000;Scientific Committee on Animal Health and Animal Welfare 2000;Thompson 1994;Travis 1995;Van Kruiningen 1999).

The recent identification of specific genes among a proportion of patients with CD is important in developing an understanding of the aetiology (Hugot et al. 2001). However, significant progress in treatment and prevention is likely to result from a better understanding of the environmental factors involved in the pathogenesis of CD. This includes microbial agents, where interactions between the mucosal surface of the gut with diet and bacteria will be fundamental. A number of potential mechanisms for the transmission of MAP from animals to humans have

been suggested, including drinking water, milk and dairy products, direct contact with animals in farms and zoos (Pavlik et al. 2000) and via aerosol spread (Hermon-Taylor et al 2000;Wendt et al. 1980).

MAP and drinking water

To date, there is no published evidence confirming the presence of MAP in drinking water in the UK. The UK Public Health Laboratory Service (PHLS) studies have failed to find MAP in UK drinking water (Lee J 2002), although this may be due to limitations in the methods of detection (Grant IR 1997;Whan et al. 2001). A survey of untreated water entering nine water treatment works (WTWs) in Northern Ireland using PCR and culture found MAP in 15 % of 192 one litre water samples (Whan 2003 cited in Grant 2005). Mishina et al in 1996 found that a previously isolated *M avium* from the municipal water supply of a major US city was in fact *M avium subspp paratuberculosis* (Mishina et al. 1996). Other mycobacteria (not MAP) have been found in the water supply in Paris, France (Le Dantec et al. 2002d).

The resilience of MAP in the environment and its presence in the faeces of cattle, with potential to contaminate surface water sources would suggest that transmission may not always be eliminated by water treatment. For example, slow sand filtration is more efficient in removing mycobacteria compared with rapid sand filtration (Le Dantec et al. 2002c). It is also known that MAP survives two parts per million chlorination after artificial inoculation of water (Whan et al 2001) and is twice as resistant as *E coli* to chlorination (Le Dantec et al. 2002b). Resistance to the effect of chlorination increases with low nutrient, low temperature and increased pH conditions that are common in water systems (Le Dantec et al. 2002a). This strongly suggests that MAP can survive the usual level of chlorination.

It is also likely that MAP can adapt within protozoa in the environment and in biofilm communities increasing its survival and possibly inducing a change in phenotype and virulence (Willett 1998). It is also theoretically possible that MAP could infiltrate into treated water supplies through breaks in the distribution network, where outside ingress is possible.

MAP and milk

MAP has been isolated from both raw (Grant et al. 2001) and pasteurised commercial (Grant 2003) milk in the UK and from commercial milk in the US (Ellingson J 2004) and Switzerland (Corti et al. 2002). Furthermore, MAP has been detected in cattle clinically (Taylor et al. 1981) and subclinically (Streeter et al. 1995) infected with Johne's disease. However, studies have shown that MAP is not eradicated from milk by standard High Temperature Short Time pasteurisation procedures (72 ºC for 15 seconds) (Grant et al. 1996). In a more recent study the same group found 10 culture positive milk samples from eight of 241 dairy processing establishments in the UK. This includes two samples that were treated at 72 ºC to 75 ºC for a longer duration of 25 seconds (Grant et al. 2002). Two comprehensive reviews on the ability of pasteurisation to eliminate MAP from milk concluded that MAP may occasionally survive commercial pasteurisation, thus providing a potential route of transmission to humans (Grant 2003;Lund et al. 2002).

A majority of the studies available have used milk spiked with laboratory grown MAP (Grant 1998;Meylan et al. 1996;Sung et al. 1998). However, unfortunately it is unclear whether laboratory strains of MAP are more or less resistant to heat treatment. Furthermore, the sensitivity of culture methods used to detect MAP is currently poor. Culture conditions may not be optimal and therefore subsequent failure to grow colonies of MAP following heat shock in conventional

cultures may not necessarily prove the absence of viable MAP. Further potential sources of error in such studies include a possible disabling effect of freeze-thaw and sonication on MAP prior to heat-shock.

In view of the possibility that MAP may contribute to the aetiology of CD together with a lack of epidemiological evidence to support or refute this hypothesis, a case control study was conducted to determine whether MAP contamination of drinking water or milk and dairy products is associated with CD.

AIM

To investigate a possible role for MAP in the aetiology of CD using a case control study. The study will determine whether individuals consuming water or milk and dairy products potentially contaminated with MAP are at a higher risk of developing CD.

OBJECTIVES

1. To determine whether drinking potentially MAP contaminated water is associated with the risk of developing CD.
2. To determine whether intake of milk and dairy products is associated with a higher risk of CD.

METHOD

Study design

A case control study was conducted in five English regions, in which proxy markers of water contamination with MAP and also dairy product and drinking water intake were compared between patients with CD and a control group. The choice of five centres from different parts of the country ensured heterogeneity of environmental exposure. The study received ethical review and approval by the Metropolitan Medical Research Ethics Committee and from local research ethics and governance committees in all study centres.

Study centres

The five centres involved in the study included:

a. Norwich: Patients were recruited through outpatient clinics in the Norfolk and Norwich University Hospital, Norwich, James Paget Hospital, Great Yarmouth and Cromer & District Hospital, Cromer. Estimated catchment population 500,000.

b. Leicester: Patients were recruited from three hospitals; Glenfield Hospital, Leicester General Hospital and Leicester Royal Infirmary. Estimated catchment population 900,000.

c. Liverpool: All patients were recruited from the Royal Liverpool Hospital - Estimated catchment population 450,000.

d. Bristol: Patients were recruited from the Bristol Royal Infirmary - Estimated catchment population 405,000.

e. Sheffield: All patients were recruited from outpatient clinics in the Royal Hallamshire Hospital - Estimated catchment population 540,000.

Study participants

Case Definition: All patients presenting with CD diagnosed within five years prior to the commencement of the study and whose date of onset of symptoms was within the same five year period were considered eligible.

Patients with a more recent diagnosis were studied because their recall bias for dietary factors is less than those diagnosed more than 10 years ago (Willett 1998). The diagnosis of CD was confirmed by recording information on diagnostic investigations (radiological, endoscopic and histological) from patients case notes using the internationally accepted diagnostic criteria formulated by Lennard-Jones (Lennard-Jones 1989) (see appendix 1). Patients under 18 years were excluded as measurement of risk factors and confounders in this group may be unreliable.

Identification of cases (CD patients)

Case recruitment took place between July 2003 and June 2004. All patients with a diagnosis of CD found to be eligible based on the case definition outlined above were approached and those consenting included. Information on date of birth, gender and address of non-consenting patients was retained to check for evidence of bias in selection of cases.

Patients with CD were recruited using two main procedures, namely clinic recruitment and computer databases. The two systems were used simultaneously because several hospitals did not

have an existing database identifying all patients diagnosed with CD. Searches on available databases were a useful addition in instances where patients were well and therefore had infrequent follow-ups. A histology database was used in Leicester because of the large catchment population, wide geographical area and potentially infrequent outpatient visits by diagnosed patients who are clinically well. In the Norwich centre, patients admitted with a diagnosis of CD were also identified in wards and included if they fulfilled study criteria. In all centres, consultants' secretaries were asked to keep a copy of clinic letters of all patients who had recently attended outpatients with CD.

The research nurse in each centre reviewed the letters and identified eligible patients. The clinical notes of patients identified by research nurses were reviewed to confirm CD and to ensure the diagnosis was made within the last five years. All eligible patients were given a study pack containing a cover letter signed by the patient's consultant, written information about the study, a consent form, questionnaire and pre-paid return envelope. Patients with a clinic appointment within six weeks of identification were approached in medical and surgical outpatient clinics and given the study packs. For other eligible patients study packs were sent via the post. A local telephone number for further information was supplied. Patients were also given the option of arranging to see the study co-ordinator.

Patients who did not return the questionnaire within three weeks were telephoned by the local research nurse to ascertain if they were interested in participating. Those still interested in taking part were sent another study pack. Those declining during the telephone call and those that did not return the second questionnaire were considered non-responders, and no further contact was made (figure 1). The general practitioners (GPs) of all patients who agreed to complete the questionnaires were informed with the patient's permission.

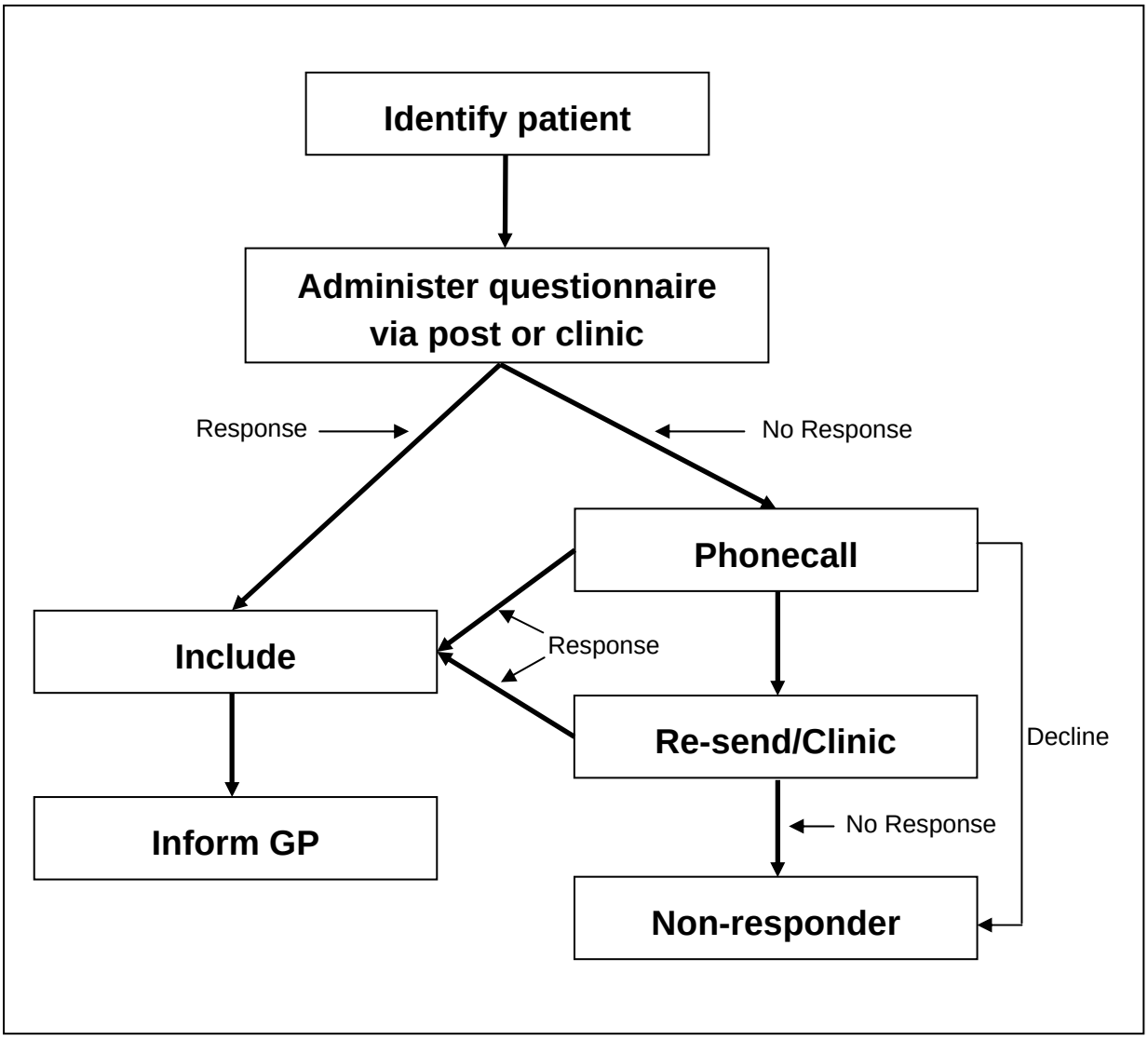

Figure 1 Identification and selection of cases

All data supplied by patients was stored on computers with password protection. Data was transcribed manually directly from the questionnaire onto the purpose-built study database. Prior to analysis, to check for input errors 10 % of the questionnaires were selected at random and dual entered by the research associate. The data was also examined for consistency prior to analysis. This included checking that, where applicable, the cases and controls were matched by the correct

date of birth range and/or gender; questions were answered in relation to the correct symptomatic and corresponding pseudo-symptomatic year; and male participants were not reported to be taking any oral contraceptives. No major discrepancies were found.

Selection & recruitment of controls

A community-based control group was used. Two methods were utilised to recruit controls. In the Norwich and Sheffield centres, controls were recruited through GPs. In all other centres the controls were chosen using the 2004 electoral register. Recruitment through GPs proved effective although time-consuming. Thus, due to time constraints, the electoral register method was employed in three of the study centres. The basic method for obtaining the sample was the same in all centres, only the sampling frame differed. In both cases multistage random sampling was utilised. This method was used because it generated a control population that was representative of the study base from which our cases arose.

In Norwich and Sheffield, two GPs per recruited case were randomly selected from a list of all GPs within the boundaries of the catchment area (i.e. those practices from which patients were referred to the Norfolk and Norwich/James Paget hospitals or Royal Hallamshire hospital). In brief, the name and postcode of each practice referring patients over the preceding 12 months to the medical gastroenterology and lower gastrointestinal surgical clinics of each hospital was obtained. Using this data, all practices were assigned to a Strategic Health Authority (SHA). This list was used to calculate the probability of referral of cases to the local hospital from each practice compared with all other practices in the same SHA area. The probability of randomly selecting a particular GP was directly proportional to the number of patients referred to the local hospital's gastroenterology service from their surgery. Therefore practices with a larger number of referrals were more likely to

be selected. This overcame the problem that hospital populations are not simple random samples of their catchment population.

Each GP identified was sent an invitation letter and written information sheet, followed a week later by a telephone call from the local research nurse. GPs were asked to consider identifying five randomly selected controls from their list. These individuals were frequency matched for age (± one year of birth) and gender of the case with which the GP was linked. An alphabetical list of all patients registered with the GP was generated and used as the sampling frame. Eligible controls were selected by the research associate or research nurse in each centre using a table of random numbers. If the GP felt unable to participate, another GP from the same practice was approached. If they also declined, a further practice was randomly selected. GPs were asked if they would consider sending a signed letter to the controls on practice notepaper. If this was not possible, hospital headed notepaper was used, signed by the consultant gastroenterologist in the local centre. Assuming a 20 % response rate, the aim was to match each case with a total of two controls, one from each of two different general practices in the hospital referral area.

In the remaining three centres, the hospital catchment area was defined using the same method and two GPs per case were randomly selected. As previously, the probability of randomly selecting a particular GP was directly proportional to the number of patients referred to the local hospital's gastroenterology service. The subsequent procedure was different. A geographical catchment area was allocated to each GP practice using the Geographical Information System (GIS) software based on the assumption that patients living around a practice would register with a GP in the same area. This assumption will be incorrect in a proportion of cases where individuals register with a GP, for example, near their place of work. Postcodes falling within each surgery catchment were identified and two postcodes within this catchment area were randomly selected. Using these two

postcodes, five controls of the same gender as the case were randomly selected from the list of individuals living within those specific postcodes and present on the electoral register. Therefore each case was matched by gender to 10 randomly selected controls. Both the postcodes and controls within each postcode area were selected by the research associate using the table of random numbers. The same study pack including a cover letter signed by the local gastroenterologist in each centre, a written information sheet, consent form, questionnaire and pre-paid return envelope, was sent to all selected controls.

In all centres, controls not responding after three weeks were sent one further invitation to participate in the study. All individuals not responding three weeks after the re-send were regarded as non-responders, and no further contact was made. The procedure is summarised in figure 2.

Improving response rates among cases and controls

Several measures to improve the response rate from cases and controls were used based on evidence identified from the published literature (Edwards et al. 2002). These included the use of a coloured questionnaire, first class postage and return mail, a minimum of two contacts, a telephone call to non-responding cases, a cover letter signed by the patient's consultant, or for controls, by their GP or a local hospital consultant.

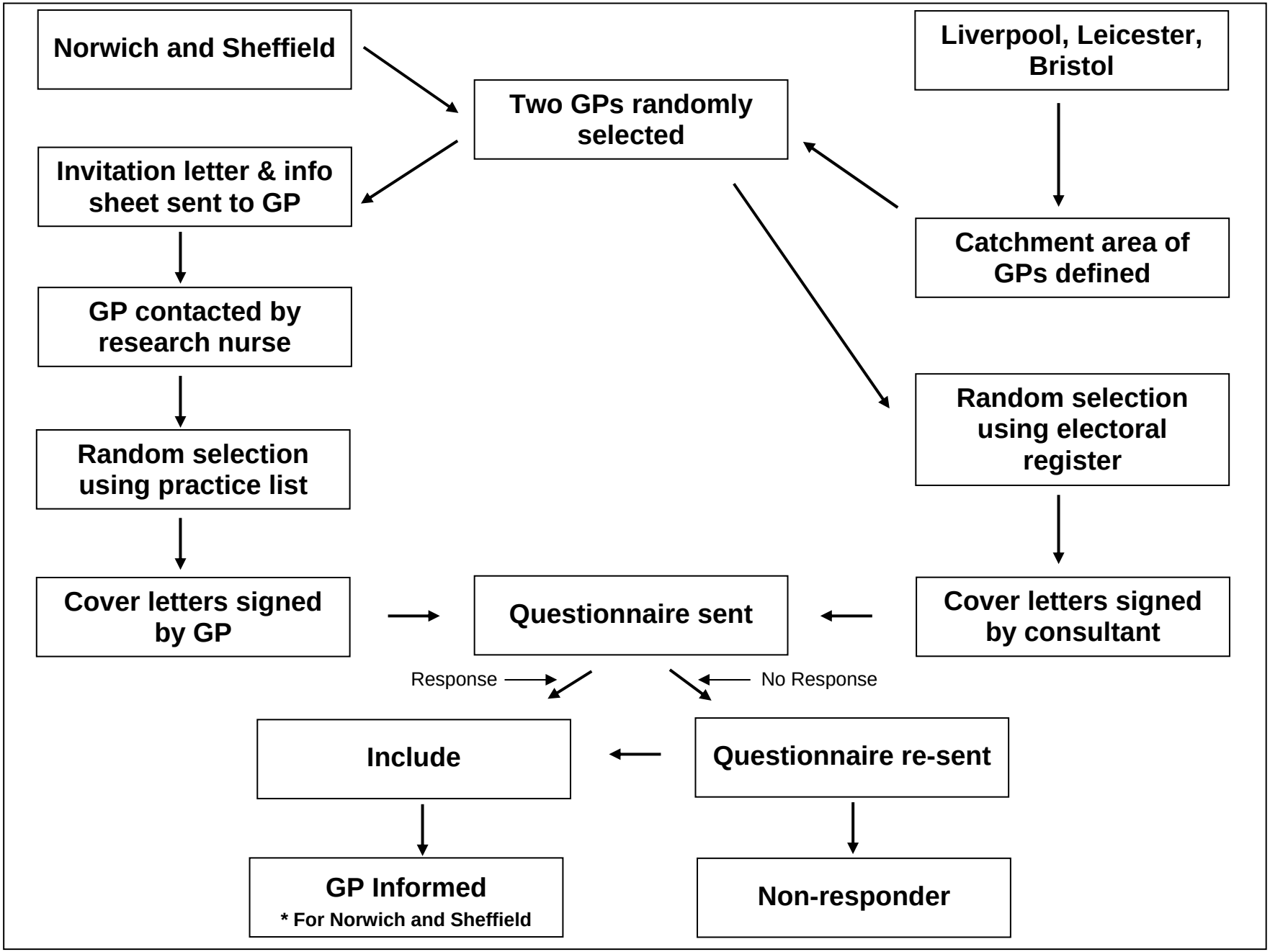

Figure 2 Identification and selection of controls

Measurement of exposures

Two main methods were used to ascertain exposure, both based on proxy measures which assume MAP from animals contaminates water and dairy product supply (figure 3).

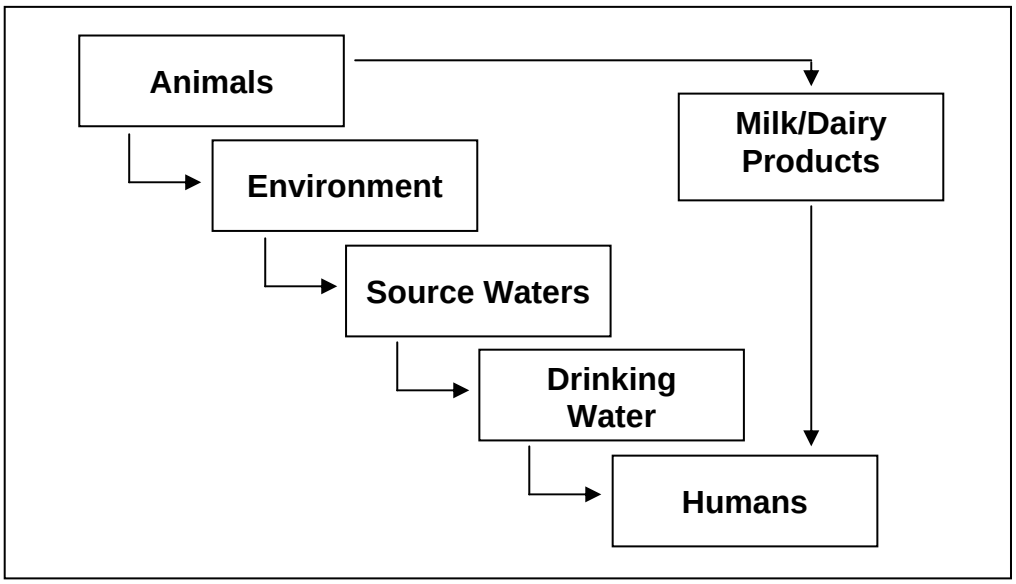

Figure 3 Potential routes of MAP exposure from animals to humans

Proxy measures of potential exposure to MAP via drinking water

In order to obtain proxy measures of potential drinking water exposure to MAP a two stage method was applied. The first involved tracing the source of each participant's water through the distribution system and ultimately to the catchment from which it was derived. The second stage involved estimating the risk of MAP at all these stages.

Tracing of a participants water supply

An outline of the procedure is illustrated in figure 4 for a hypothetical participant and can be divided into three distinct stages:

1. Individual to WTWs
2. WTWs to abstraction point
3. Abstraction point to catchment

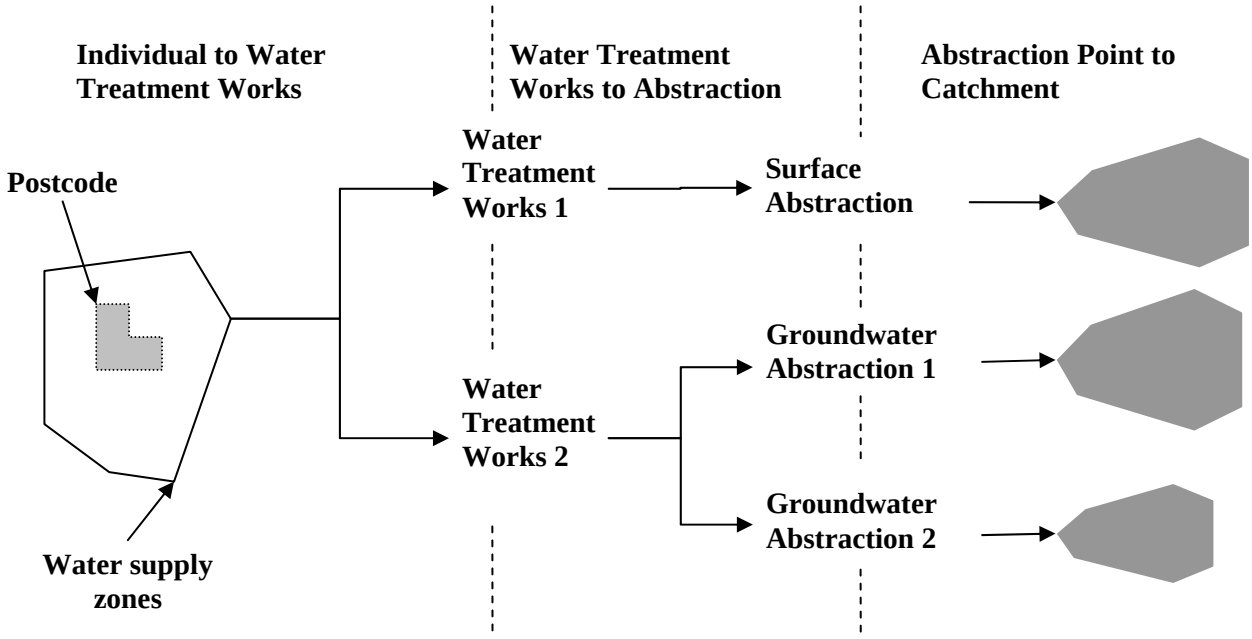

Figure 4 Tracing a respondent's water supply

1) Individual to WTWs

The postcode of the participants' current and previous residential addresses (specifically in 2000, 1998 and 1993) as well as their current and previous place of work were obtained from the questionnaire. It was assumed that these were the points where public water supplies were consumed. Each of these postcodes was converted to a grid by cross referencing with the Ordnance Survey Code-Point database which provides a grid reference for each of the 1.7 million unit postcodes in the UK. However, several postcodes did not have an entry on Code-Point, the most likely reason being that they are no longer in use by the Post Office. In such circumstances, these postcodes were additionally cross referenced with the All Field Postcode Directory which contains historical postcode data.

Once grid references had been obtained, the Water Supply Zone (WSZ) associated with these were identified. WSZs delineate geographical areas where supplies are from an identical source or sources, and usually serve less than 100,000 individuals. The digital boundaries of all the WSZs in England and Wales were obtained from the Drinking Water Inspectorate (DWI). Approximately 1 % (66 cases and controls) were excluded from the analyses as water supply zone data was not available for the appropriate period. Using GIS, a point in polygon procedure was used to identify the WSZ associated with each of the grid references.

The WTWs supplying the WSZ were then determined. In some circumstances, this may be more than one, as illustrated in figure 4, with WTW1 and WTW2 both supplying water to the WSZ. Using the details file supplied to the DWI by each water company (Water Supply (Water Quality) regulations 1989) the proportion of water supplied to each WSZ from different WTWs was identified. During this process respondents were excluded if the details files were incomplete.

2) WTWs to abstraction point

For each WTWs the abstraction points were identified. The only comprehensive data on these were collected under the Water Supply (Water Quality) (Amendment) Regulations 1999. This required all water companies to undertake a risk assessment of their water supply systems for *Cryptosporidium* and part of this involved listing the abstraction points supplying each WTWs. Where more than one abstraction point was present, the relative use of each source was estimated from the Environment Agency's Abstraction Licensing Database. This is illustrated in figure 4, where all the water for WTW1 is supplied by one surface abstraction but WTW2 is supplied by two groundwater abstraction points.

3) Abstraction point to catchment

A grid reference for each abstraction point was obtained by cross referencing the name given on the *Cryptosporidium* risk assessments with the name on the Environment Agency's national Abstraction Licensing Database. The latter contained a grid reference for each abstraction point. For surface waters a digital map of land heights was obtained from the NERC National Water Archive (Hydrological Digital Terrain Model) and these land heights were used to calculate the catchment of each surface abstraction point using GIS. This process is illustrated in figure 5 for the Heigham WTWs in Norwich.

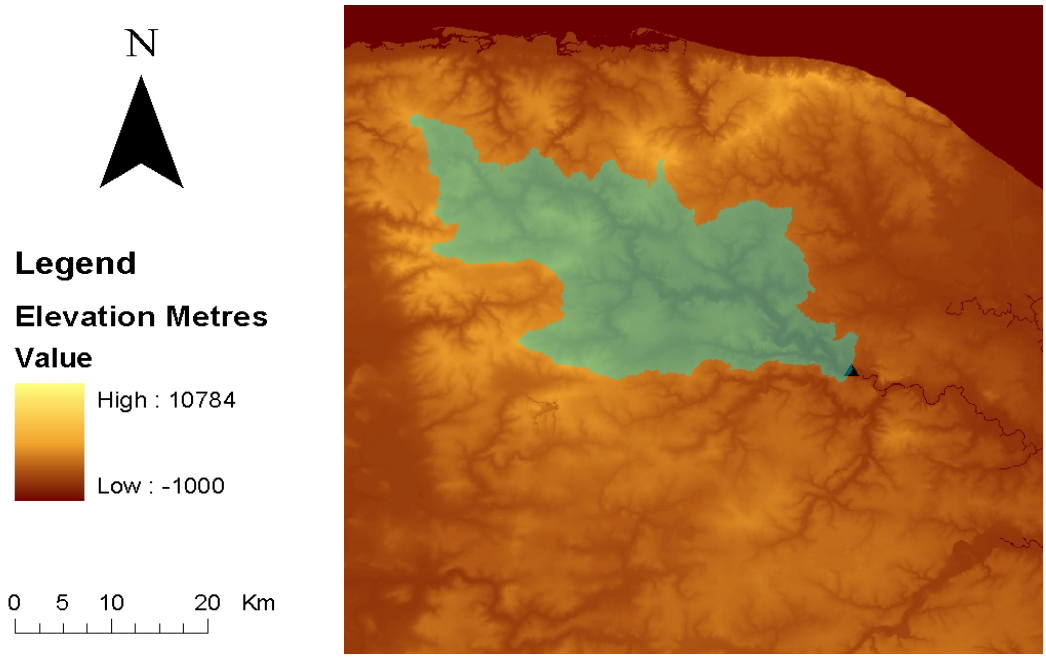

Figure 5 The catchment of Heigham WTWs. The elevation is displayed in brown with lighter colours indicating higher elevations. The catchment is shaded in green.

Groundwater catchments were obtained from the borehole source protection zone created by the Environment Agency. These combine information on the borehole abstraction rate with the hydrogeological characteristics of the aquifer to estimate a set of catchments for each groundwater abstraction point, based upon the time it takes water to move to the abstraction point. For this study the outer protection zone was used for the catchment, which equates to a 400 day time-of-travel-zone.

Potential risk factors

Once the source of each participant's water had been traced through the distribution system and ultimately to the catchment from which it was derived, proxy measures of the potential for MAP contamination along this pathway could be determined. The measures identified were:

1. Biofilm - Potential for biofilm build up in the pipes supplying water to each respondent.
2. Water treatment – Information on the level of treatment to which each respondent's water was subjected.
3. Water source – Whether the supply of water was river or groundwater.
4. Livestock density in water catchment area - Density of all livestock in the geographical catchment of the WTWs.
5. Groundwater protection – The protection of groundwaters from surface water infiltration.
6. Johne's disease – The number of reports of diagnoses of Johne's disease in symptomatic animals to the Veterinary Laboratories Agency in England between 2000 and 2002 by county.

1) Biofilm

The potential for biofilm build up in the pipes supplying water to households was estimated by calculating the straight line distance (km) between each participant's postcode and the WTWs supplying the water. This measure is a proxy for the length of water pipes with the underlying hypothesis that the greater the pipe distance the greater the potential for biofilm contamination of the water.

2) Water treatment

Data on the types of treatment in use in the WTWs were derived from the aforementioned *Cryptosporidium* risk assessments. These data were obtained from the DWI, entered into a database and queried to identify the level of water treatment for each of the WTWs supplying the respondents. This identified the use of sedimentation, coagulation, flocculation or filtration. It was hypothesised that each of these treatments may vary in the efficacy with which they will remove MAP.

3) Water source

For each WTW the source of the drinking water (surface, groundwater or mixed) was identified from the site details file supplied to the DWI under the Water Supply (Water Quality) regulations 1989. The selection of water source as a proxy marker was based on the assumption that MAP contamination of water supplies derived from surface sources is more likely.

4) Livestock density

To estimate the risk of MAP in the catchment the livestock density was calculated. These data were obtained from the DEFRA agricultural census supplied by the University of Edinburgh Data Library. This source has combined agricultural census data with land use data to estimate agricultural activities on a 2 km^2 cell. Using the GIS the average and maximum density (cattle /km^2) of total cattle; dairy and beef cattle, sheep and pigs, were calculated for all the surface and groundwater catchments. This is illustrated in figure 6 which displays the catchment for Heigham WTWs combined with information on cattle density.

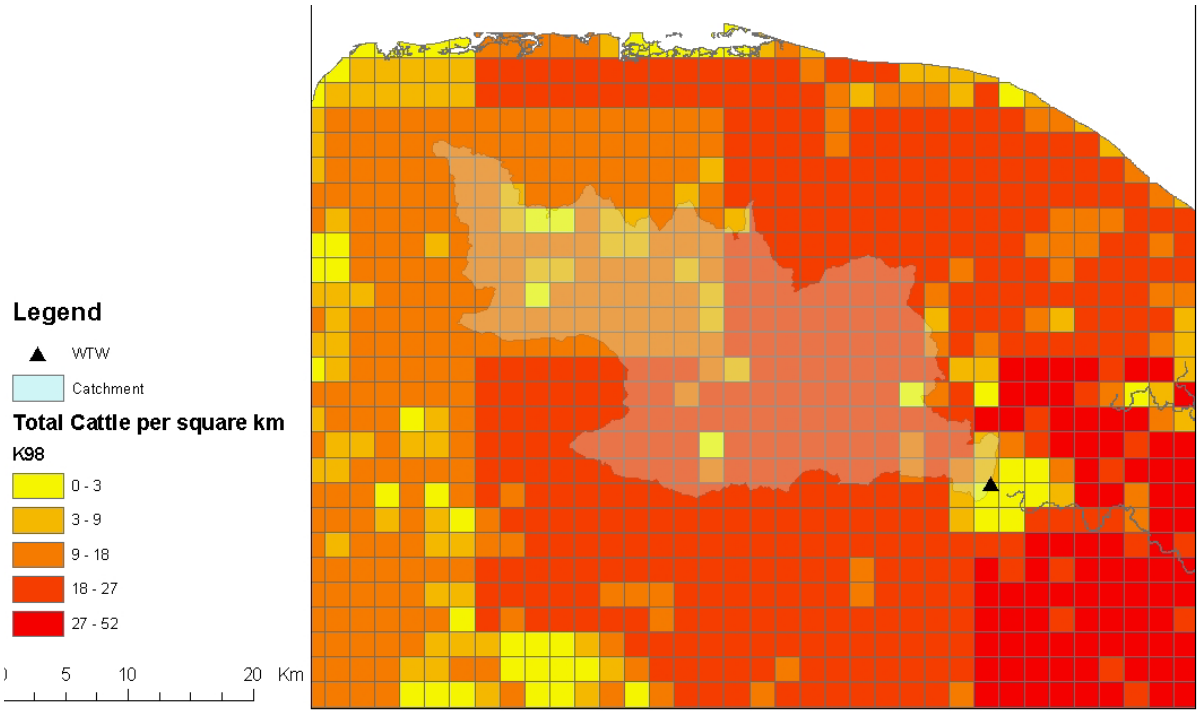

Figure 6 Water catchment risks calculated by combining catchment data from Heigham WTWs with the cattle density data for this area.

5) Groundwater protection

When considering groundwater catchments it is imperative to consider the nature of the aquifer as a variety of factors may alter the potential for MAP to pass through to the water. For example, a shallow aquifer is more likely to be contaminated with MAP than a deep one. Therefore, the depth of each aquifer was estimated using an Environment Agency database of 3,714 borehole depths and identifying the nearest borehole to the abstraction point from this database. In addition, for each catchment the Environment Agency Groundwater Vulnerability Maps were used to estimate the degree of protection afforded to the aquifer by the aquifer properties, the presence of any overlaying drift material and the characteristics of the soil. This was achieved

using a point in polygon procedure to identify the soil, drift and aquifer properties at the abstraction point.

6) Johne's disease

A measure of the occurrence of Johne's disease in cattle was estimated by dividing the number of reports of Johne's disease between 2000 and 2003 in each county, obtained from surveillance data provided by the Veterinary Laboratories Agency, by the total number of cattle in each area from the annual Cattle Census (DEFRA).

Exposure data obtained from questionnaires

For cases, all questions were asked in relation to the year prior to the development of symptoms (between 1998 and 2004). For the purposes of the study, all controls were assigned a pseudo-symptom date – this was the year in which the frequency matched case was last asymptomatic. Controls were asked to complete the same questionnaire in the period corresponding to the year of the pseudo-symptom date (see appendix 2). For both groups this time period will correspond to the 'year of interest' throughout this report.

Questionnaire measurement of water type and quantity consumed

All participants were asked questions on past water intake, including the quantity of water drunk both at home and work, and their postal addresses over the last 10 years. With this information, the GIS was used to calculate the proxy markers of MAP exposure outlined previously.

Addresses over the last 10 years were collated since it is currently unknown how many years prior to the development of symptoms, patients' exposure to MAP occurs.

Measurement of dairy product intake

Dairy product intake was measured using questions based on those which have been used and validated in large cohort studies, namely the Nurses' Health Study (Willett et al. 1987) and EPIC (European Prospective Investigation Into Cancer and Nutrition) (Bingham et al. 1997;Day et al. 2001). This design asks participants to record the frequency of dairy product consumption according to nine standard categories (none/< 1 per month, 1-3 per month, 1 per week, 2-4 per week, 5-6 per week, 1 per day, 2-3 per day, 4-5 per day, 6^+ per day). For most food items, a standard measure is described e.g. glass of milk, carton of yoghurt. Several other food questions on non-dairy products were also included, such as fish, meat (beef and canned meat products) and certain fruits and vegetables (apples, bananas & peas). There are no known associations with vegetables and red meat in CD. These "dummy" questions were included to assess whether patients think all foods listed are associated with an increased risk of CD. Questions on fish oil were considered relevant since limited trial work has shown that this may prevent intestinal inflammation by reducing the production of pro-inflammatory mediators such as prostaglandin E_2 and leukotriene B_4 (Belluzzi et al. 1996;Hawthorne et al. 1992;Stenson et al. 1992). Studies have shown that past diet can be recalled with reasonable accuracy up to 10 years previously (Willett 1998).

A finding in cases of a raised dairy product intake, low fish consumption and normal intake of fruits etc, would help support the validity of the dairy product data. Cases were asked to recall their dietary habits in the year before symptoms developed (year of interest). Controls were also asked to

complete the same questionnaire which asks about past diet, corresponding to the number of years ago that the matched case was last asymptomatic (see appendix 2, questions 16-19).

The intake of all water and food items per individual were estimated from the food frequency data collected by calculating the amount consumed in a 30 day period. This quantity was multiplied by volume (mls) or weight (grams) of the standard portion sizes stated in the questionnaire to obtain the amount per 30 days. This method has been previously validated for the UK population. Each food variable was then categorised into five quintiles.

Measurement of confounders

Smoking (Calkins 1989;Vessey et al. 1986), family history (Logan et al. 1989) and use of the oral contraceptive pill (Vessey et al 1986) are associated with an increased risk of CD and were measured using the questionnaire (see appendix 2, questions 20-25). Although the mechanism of action is unknown it is possible both smoking and oral contraceptive use could provoke mesenteric infarction leading to intestinal inflammation (Hudson et al. 1996). Several other variables were collected to explore alternative factors that are thought to increase the risk of CD or affect other observed associations. These factors included appendectomy (Ekbom et al. 2004;Loftus, Jr. 2004), farm visits, farm holidays, dish washing and awareness of the causes of CD. For further details refer to the study questionnaire in appendix 2.

Sample size calculations

The sample size was determined based on five study centres (Norfolk, Leicestershire, Liverpool, Sheffield and Bristol) with an estimated total catchment population of 3 000 000. Assuming the incidence of CD is 5/100 000/year (Logan et al 1989) with 80 % of patients aged over 18 years and

70 % of these agreeing to participate, a minimum of 104 CD patients were required. This would provide a power of 80 % to detect a doubling of the risk of CD associated with a surface water supply if cases were matched to two controls from either the GP list or the electoral register. It was assumed that 60 % of controls have a ground water source.

Analysis

The aim of the analysis was to demonstrate whether the risk of CD is greater in individuals consuming significant quantities of water from areas where the drinking water supply is at risk from MAP contamination. Further analysis was used to assess how dairy product intake may affect this risk and control the effect of other known confounding factors.

Data entry

The research associate and a trained research assistant entered study data into a purpose built Microsoft Access database. The database consists of three sections, the study questionnaire, the clinical diagnostic criteria information and data on proxy measures of exposure generated using the GIS. Where addresses were supplied without the relevant postcode, the Royal Mail website was used to ascertain the postcode for entry into the database.

Descriptive analysis

Characteristics of cases and controls

The frequencies and percentages of the variables among cases and controls were determined. Simple associations were assessed with frequency tables and Pearson's chi square tests or the

exact test where appropriate for two independent proportions. Means and standard deviations were calculated to summarise continuous effects and were compared by t tests or appropriate non-parametric tests when distributional assumptions were in doubt. When data have been categorised by quintiles, these were based on the overall data.

Diagnostic criteria

The information collected based on the Lennard-Jones diagnostic criteria were tabulated to identify the types of CD among the cases included in this study. The region of the gastrointestinal tract (GIT) affected was also tabulated. The type and location of CD were then categorised according to the Vienna classification of CD adapted by Gasche et al (see appendix 3)(Gasche C et al. 2000).

Logistic regression: Univariable and Multivariable analyses

Analysis of questionnaire data from all cases with frequency matched responding controls

The analysis included all responding cases and all responding controls. Odd ratios (ORs) and 95 % Confidence Intervals (CI) were calculated comparing the proportion of cases and controls receiving water supply from areas at different levels of MAP risk. The ORs were calculated using logistic regression. The risk of MAP in the water supply and intake of dietary factors were divided into quintiles and unadjusted and adjusted ORs were calculated.

Although the original protocol proposed a matched analysis, the findings presented are for models fitted with unconditional logistic regression, adjusted for the matching criteria. Frequency matching rather than perfect individual matching was used. Unmatched analyses were used as age matching of controls with cases was not possible for the three centres using the electoral register

procedure. Undertaking a strict matched analysis would require the exclusion of many participants. Moreover, when data on a risk factor were missing for a case or control, the entire pair would be excluded from all analyses. Therefore, the age-matching criteria was ignored and frequency matching of cases and controls by gender was used.

The study data were re-analysed using conditional logistic regression matched by gender and the results were not shown to be different from the unmatched findings. In addition, for centres where age matching was possible results from matched and unmatched analysis were similar, hence we have presented the unmatched results.

To check for evidence of a non-linear association, continuous variables were recoded into equal sized categories. Odds ratios for categories were examined and no evidence of a non-linear association was found. Therefore, the models assuming a linear association were presented.

Analysis of GIS generated exposure data

Proxy measures of exposure were analysed using the postcode reported by cases and controls for the year of interest (year prior to the development of symptoms or pseudo-symptom date), three years (2000), five years (1998) and 10 years (2003) prior to the start of the study. The results were initially tabulated and examined for association between CD and each of the proxy measures of exposure (water source and treatment) using chi squared tests and t test for mean, where appropriate. The specific proxy measures available for comparison include type of water treatment (sedimentation, coagulation, flocculation and filtration), type of water source (classified as ground, surface or mixed water), measures of ground protection (geological class, presence of drift material, borehole depth), animal density (cattle, sheep and pigs) and the number of reports of Johne's disease in each county. All supplies were disinfected and so no analyses were done on this variable.

Logistic regression models were used to calculate the OR for each proxy measure of exposure to MAP contaminated water. Univariable ORs were calculated for the variables identified above, including measures of water source and treatment, animal density, ground protection and reports of Johne's disease. To account for multiple WTWs supplying water to each participant, the effect of each exposure factor was weighted by proportion of water supply from each WTWs for that individual. Similarly, to account for multiple abstraction areas supplying water to each WTWs, variables derived from the abstraction areas (cattle density, ground protection and Johne's disease reports) were weighted by the proportion of water from each area.

A total of 17 variables, potentially associated with water-related transmission, were available for calculation of ORs and 95 % CI, therefore using Bonferroni principle, a p value of 0.005 was required before variables were considered statistically significant to account for multiple hypothesis testing. Evidence of collinearity between variables was also examined. Stepwise multivariable logistic regression analysis with backward elimination was used, but there was a high level of collinearity and a proportion of the water treatment variables were incomplete. Multivariable models controlling for the effect of age, centre, smoking and family history were fitted for each variable.

Confounding

The confounding effect of known risk factors for CD such as smoking, family history and oral contraceptive use were controlled for using multivariable logistic regression models. The potential confounding effect of other risk factors such as age, gender, previous appendectomy and detergent use were also analysed.

Interaction

Effect modification in the observed association for all significant variables was analysed using likelihood ratio tests comparing a more saturated to a less saturated model. Evidence of effect modification by age on the observed significant associations between the risk factors and CD was examined.

Sub group analysis

It has been hypothesised that CD may not be one disease but rather consists of several diseases (Gasche et al. 2005;Korelitz et al. 1996). To investigate whether the different types of CD (stricturing, penetrating and non-stricturing non-penetrating) and disease affecting different parts of the GIT (upper gastrointestinal (GI), terminal ileum, ileocolon and colon) have different risk factors, univariable and multivariable logistic regression analysis was used to calculate ORs for risk factors by sub group.

All statistical analyses were performed using STATA Version 8.0 statistical package.

RESULTS

The study population included cases with symptoms and a diagnosis of CD at any time between December 1998 and December 2003. The number of potential cases and controls identified, excluded and included across all five centres is shown in figure 7. Overall 373 cases and 2,149 potential controls were identified over the duration of the study period. In total, 344 cases were considered eligible to be included in the study, and of these 218 responded to the questionnaire. For the controls, of the 2,149 individuals approached, 34 were later excluded from the analysis for various reasons (see figure 7). A total of 812 controls responded to the questionnaire. A further 19 control questionnaires were received during/after analysis of the data and were therefore not included in the study. The total number of cases and controls recruited exceeded that required in the sample size calculations. Initially the response rate was poor and therefore further measures were implemented to improve compliance which resulted in higher response figures. This in turn permitted a subgroup analysis to be conducted. The geographical distribution during the year of interest[1] of all the cases and controls who responded to the questionnaire by centre is illustrated in figures 8a to 8c.

The demographic characteristics of the study population are shown in table 1. The mean age of the cases (42 years, SD 19.3) were significantly different compared with the controls (51 years, SD 18.6), although the age range between the two groups was comparable (table 1), with a bimodal distribution seen for the cases (figure 9). In total, 57 % of the cases compared with 62 % of the control population were female (table 1). A majority of the cases who responded to the questionnaire were based in Norwich (31.2 %) and Leicester (25.2 %) with a comparable distribution between the remaining three centres. In comparison, a majority of the control population included in the study were located in the Norwich area (44.2 %) with comparable

[1] Year prior to the development of symptoms for the cases and the pseudo-symptom year for the controls

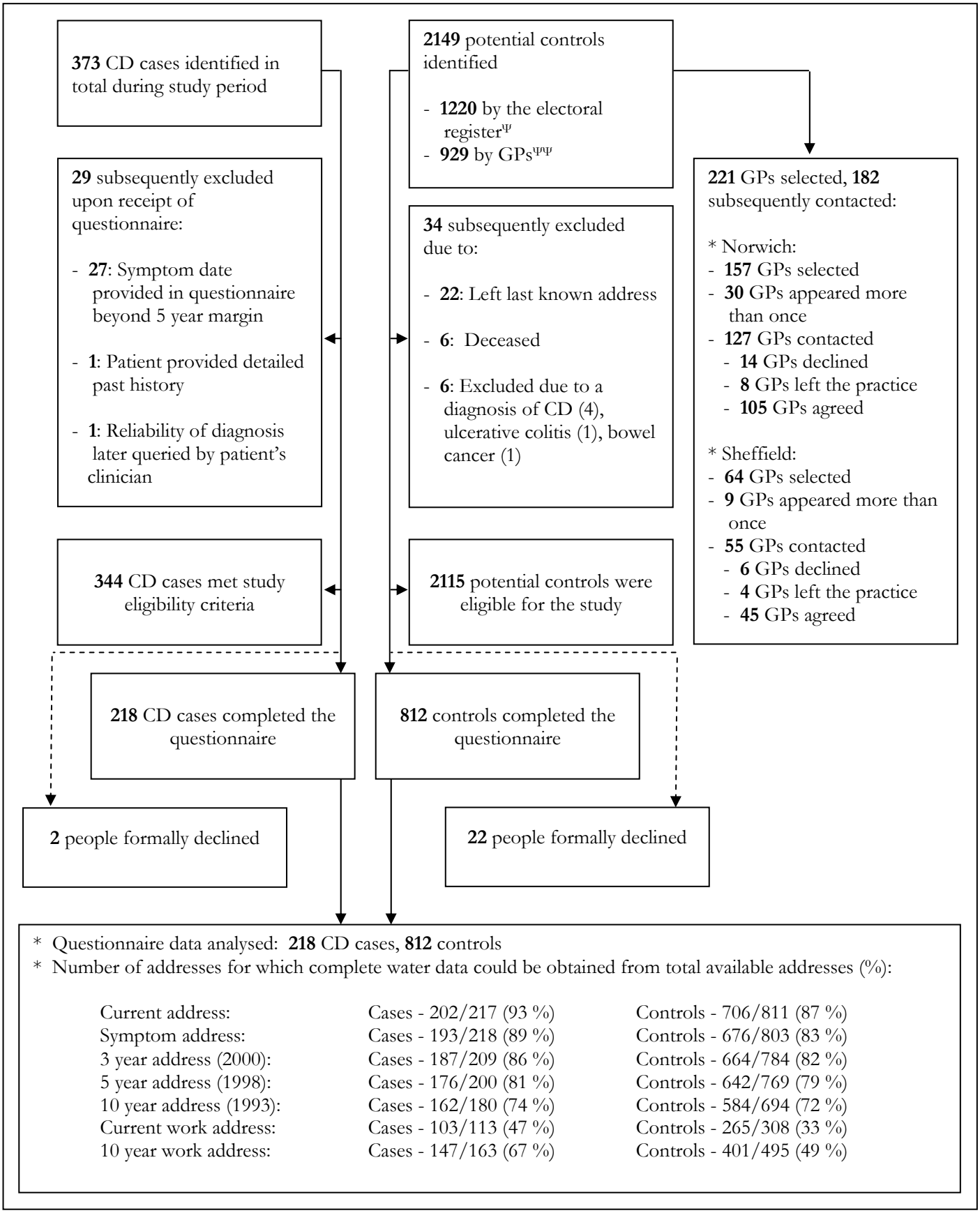

Figure 7 Number of potential cases and controls identified, excluded and included across the five study centres (Ψ Liverpool, Leicester and Bristol centres; ΨΨ Norwich and Sheffield centres)

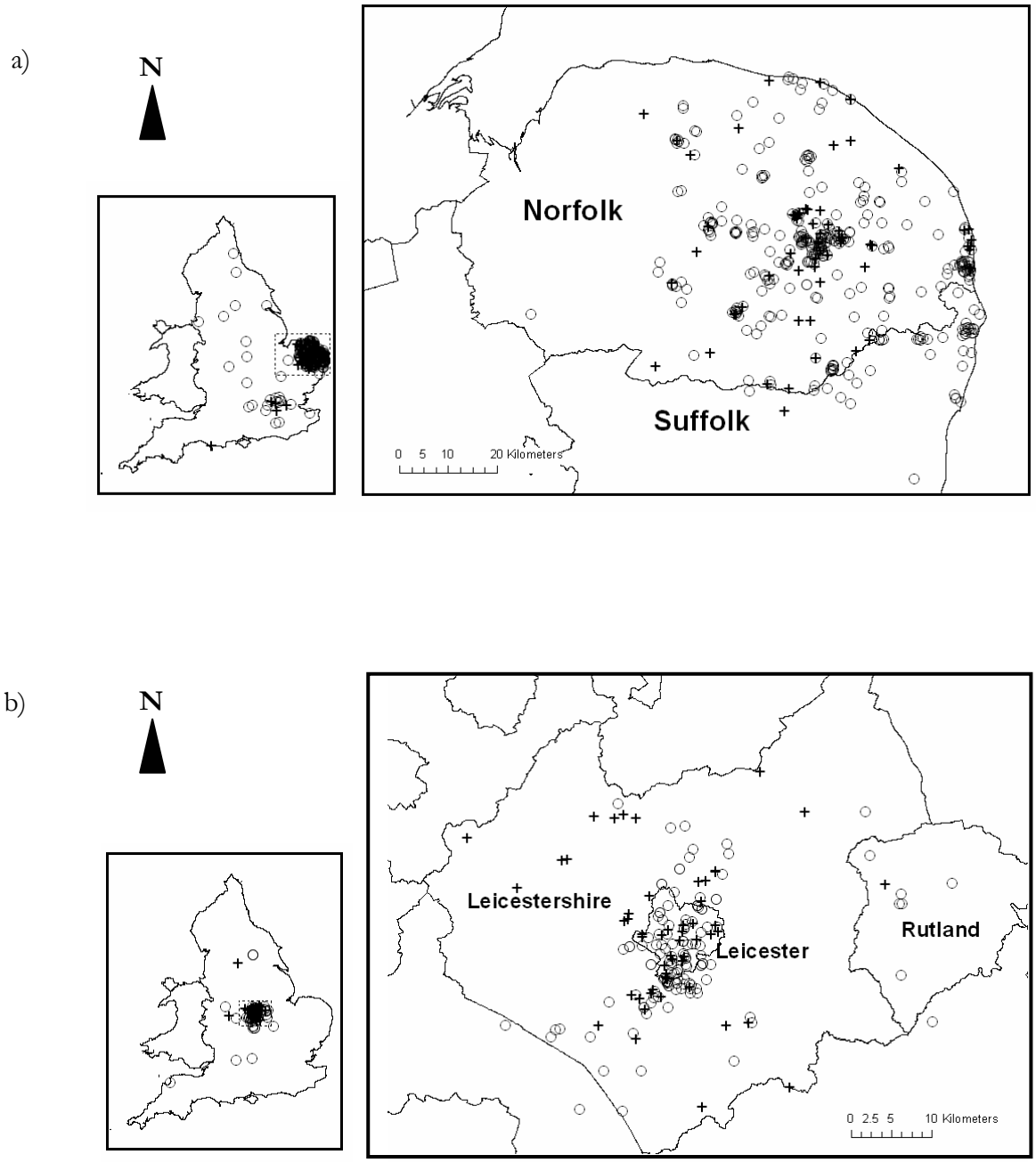

Figure 8a Geographical distribution of cases (+) and controls (O) in (a) Norwich and (b) Leicester during the year of interest.

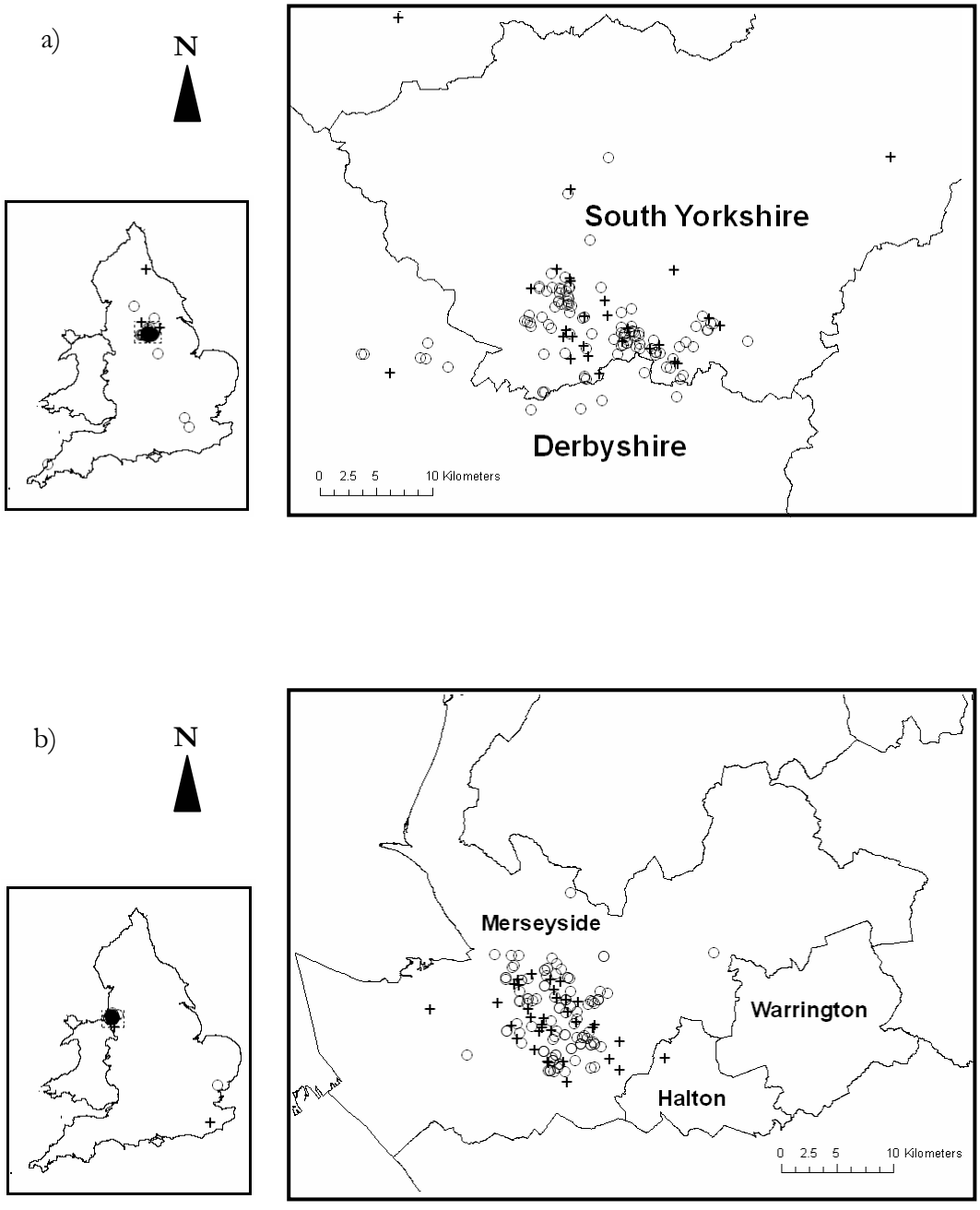

Figure 8b Geographical distribution of cases (+) and controls (O) in (a) Sheffield and (b) Liverpool during the year of interest.

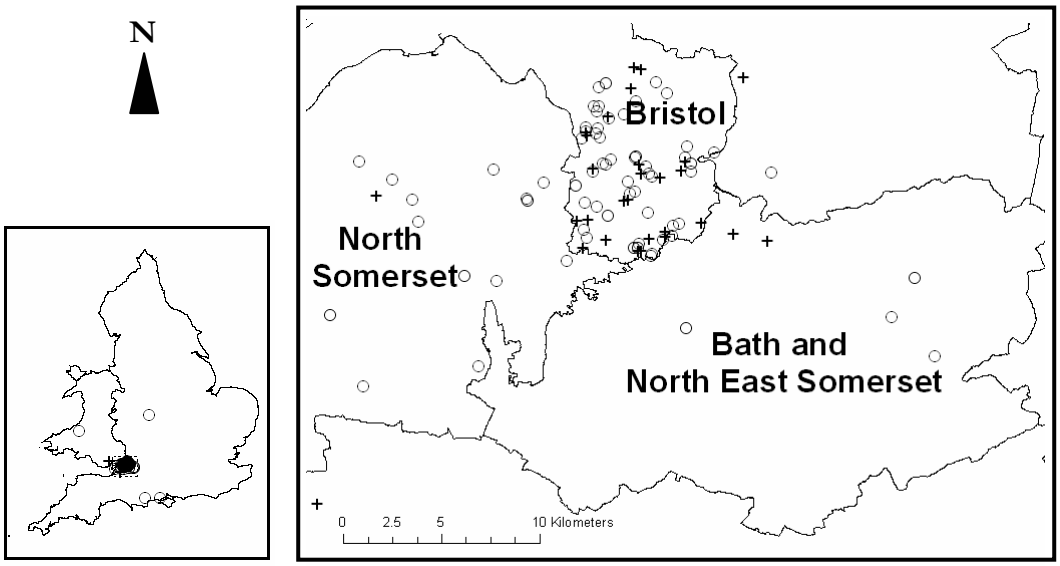

Figure 8c Geographical distribution of cases (+) and controls (O) in Bristol during the year of interest.

distributions between the other four centres (table 1; figures 8a to 8c). For both study populations, the majority of cases (92.0 %) and controls (92.5 %) lived in a house/flat/apartment during the year of interest (table 1).

Table 1 Characteristics of the study population. Values are shown as totals (%) of the group, unless otherwise stated.

Variable		Cases (n=218)	Controls (n=812)	P value*
Age (yrs)	Mean (SD)	42 (19.3)	51 (18.6)	0.00001
	Range	18 - 86	17 - 94	
Gender	Females	125/218 (57.3)	507/812 (62.4)	0.17
Centre	Norwich	68/218 (31.2)	359/812 (44.2)	0.004
	Leicester	55/218 (25.2)	146/812 (18.0)	
	Sheffield	28/218 (12.8)	112/812 (13.8)	
	Liverpool	36/218 (16.5)	109/812 (13.4)	
	Bristol	31/218 (14.2)	86/812 (10.6)	
Residence	H/F/A	196/213 (92.0)	745/805 (92.5)	0.023
	Hostel	0/213 (0)	1/805 (0.1)	
	Residential home	4/213 (1.9)	16/805 (2.0)	
	Boarding school	1/213 (0.5)	0/805 (0)	
	University	4/213 (1.9)	2/805 (0.2)	
	Other	8/213 (3.8)	41/805 (5.1)	
Disease type	Stricturing	63/217 (29.0)	-	
	Penetrating	29/217 (13.4)	-	
	NSNP	125/217 (57.6)	-	
Disease site	Upper GI	27/217 (12.4)	-	
	Terminal ileum	14/217 (6.5)	-	
	Ileocolon	95/217 (43.8)	-	
	Colon	73/217 (33.6)	-	
	Uncategorised¥	8/217 (3.7)	-	

* P value using chi square test, t test and equivalent non parametric alternatives where appropriate
H/F/A, House/Flat/Apartment; NSNP, Non-Stricturing Non-Penetrating.
¥ Cases fall into more than one group.

The type and site of CD was categorised according to the Vienna classification (Gasche C et al 2000) (see appendix 3). The majority of cases were diagnosed with non-stricturing non-penetrating CD (57.6 %), with the most common locations of disease sited at the ileocolon (43.8 %) and colon (33.6 %).

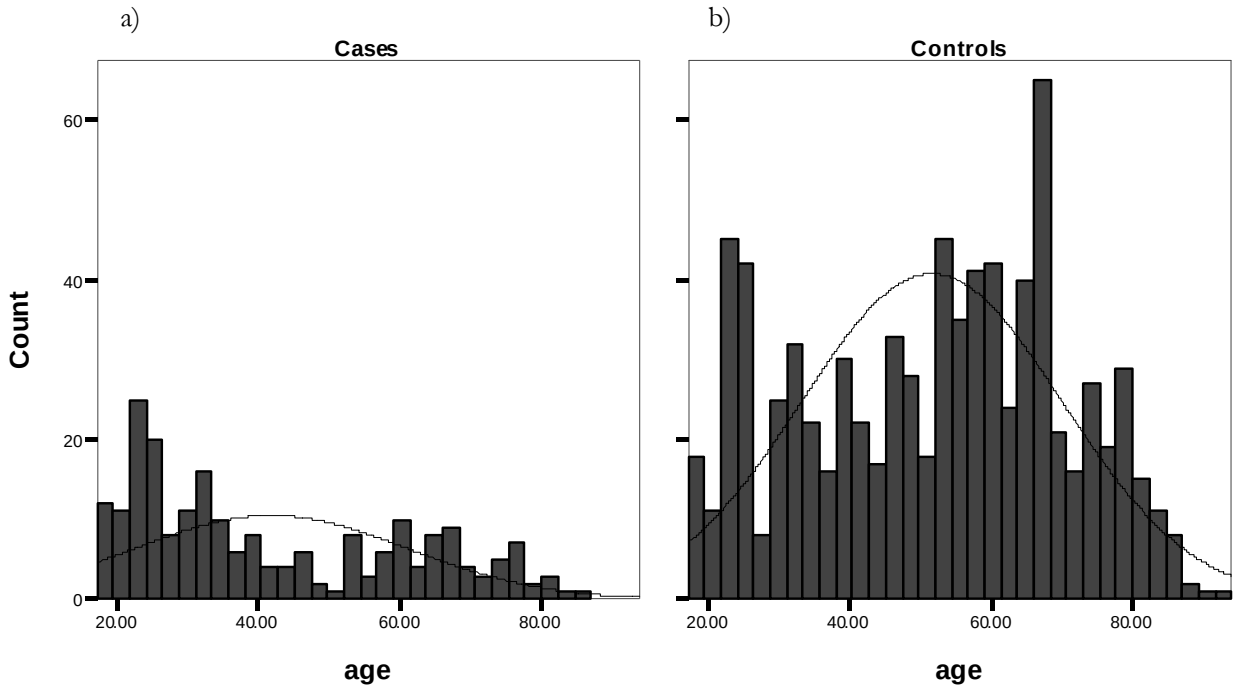

Figure 9 Age distribution of (a) cases and (b) controls

Table 2 shows the distribution of cases and controls by risk factors and proxy measures of exposure to MAP obtained using the study questionnaire. For drinking water, one of the primary factors under investigation, the mean intake of unboiled water was significantly higher among cases compared with controls (p=0.0032), however the converse was true for boiled water, although this was not significant (p=0.2092). The intake of bottled water did not differ between the two groups. A majority of both cases (98.6 %) and controls (96.9 %) received a public water supply (table 2).

Table 2 Distribution of risk factors. Values are shown as totals (%) of the group, unless otherwise stated.

Variable		Cases (n=218)	Controls (n=812)	P value*
OCP use	Yes	76/124 (61.3)	278/505 (55.0)	0.209
Ever smoke	Yes	123/211 (58.3)	340/783 (43.4)	0.0001
Smoke in YOI	Yes	95/213 (44.6)	201/801 (25.1)	0.0001
Smoking frequency	None	124/218 (56.9)	615/811 (75.8)	0.0001
	1 – 9	25/218 (11.5)	66/811 (8.1)	
	10 – 19	40/218 (18.3)	69/811 (8.5)	
	20 +	29/218 (13.3)	61/811 (7.6)	
Appendectomy	Yes	27/214 (12.6)	95/794 (12)	0.795
Family history	Yes	35/190 (18.4)	33/671(4.9)	0.0001
Wash dishes	Machine	46/216 (21.3)	197/798 (24.7)	0.300
	Hand wash	170/216 (78.7)	601/798 (75.3)	
Water supply	Private	3/216 (1.4)	15/802 (1.9)	0.634
	Public	213/216 (98.6)	787/802 (98.1)	
Holiday abroad	Yes	126/212 (59.4)	529/803 (65.9)	0.081
Farm holiday	Yes	8/213 (3.8)	29/802 (3.6)	0.923
Length farm	None	210/218 (96.3)	787/810 (97.2)	0.499
holiday	< 1 week	5/218 (2.3)	10/810 (1.2)	
	1 week or more	3/218 (1.4)	13/810 (1.6)	
Farm visit	Yes	35/201 (17.4)	129/752 (17.1)	0.931
Frequency farm visit	None	186/217 (85.7)	697/811 (85.9)	0.462
(per year)	1	8/217 (3.7)	18/811 (2.2)	
	2	2/217 (0.9)	19/811 (2.3)	
	3 – 364	18/217 (8.3)	61/811 (7.5)	
	365 and over	3/217 (1.4)	16/811 (2.0)	
Contact farm animals	Yes	24/74 (32.4)	89/327 (27.2)	0.368
Boiled water (ml)	Mean (SD) Ψ	26239 (17342)	28444 (16379)	0.2092
Unboiled water (ml)	Mean (SD) Ψ	14493 (13669)	12186 (12786)	0.0032
Bottled water (ml)	Mean (SD) Ψ	2405 (5542)	2732 (6467)	0.2465
Pasteurised milk (ml)	Mean (SD) Ψ	22767 (16592)	26009 (15813)	0.0167
Fruit (g)	Mean (SD) Ψ	6633 (5993)	8119 (6540)	0.0000
Dairy products (g)	Mean (SD) Ψ	3399 (2135)	3548 (2410)	0.3109
Meat (g)	Mean (SD) Ψ	2142 (3886)	1497 (2813)	0.0092
Fish (g)	Mean (SD) Ψ	1596 (1209)	1516 (1116)	0.9141
Cod liver oil (dose)	Mean (SD) Ψ	5 (12)	8 (17)	0.0029

* P value using chi square test, t test and equivalent non parametric alternatives where appropriate
OCP, Oral contraceptive pill; YOI, Year of interest
Ψ Frequency of intake x quantity over a 30 day period

Table 2 also outlines the average intake of milk and dairy products as measured by the questionnaire. The mean intake of pasteurised milk was significantly higher among controls compared with cases, whilst the intake of other dairy products was not different between the two groups.

For other factors obtained from questionnaires, cases consumed more meat on average than controls (p=0.0092), whereas the mean intake of fruit (p=0.0000) and cod liver oil (p=0.0029) was significantly higher among controls compared with cases. There was no difference in the average consumption of fish.

Smoking behaviour differed significantly (p=0.0001) between the two groups (table 2). The percentage of individuals who reported ever smoking (58.3 %) was greater among cases than controls (43.4 %). Furthermore, there was evidence of significantly higher rates of smoking in the year of interest for cases compared with controls (44.6 % versus 25.1 %) with statistically significant evidence of a dose response relationship (p=0.0001). Table 2 also shows that a family history of CD was significantly (p=0.0001) higher among the cases (18.4 %) than the controls (4.9 %). Other variables compared such as use of oral contraceptives, visits to farms, farm holidays, holidays abroad and previous appendectomy did not differ significantly.

Table 3 shows the dietary data categorised in quintiles of intake with one representing the lowest level of intake and a score of five for the highest level. The breakdown and average quantities of each food and water group can be found in appendix 4. The significant differences in food and water consumption between cases and controls using quintiles of intake were similar to the comparisons of mean intake, with the exception of fish, where there was no difference in mean intake but a significant difference in quintiles of intake (p=0.04; table 3).

Table 3 Distribution of food and water variables for cases and controls. Values shown as totals (%) of each group in quintiles.

Food or Drink	Quintile[$]	Cases (n=218)	Controls (n=812)	P value*
Boiled drinking water	1	61 (30.8)	197 (26.7)	0.838
	2	40 (20.2)	160 (21.7)	
	3	25 (12.8)	92 (12.5)	
	4	51 (25.8)	201 (27.3)	
	5	21 (10.6)	87 (11.8)	
Unboiled water	1	28 (14.1)	170 (23.1)	0.042
	2	56 (28.3)	214 (29.0)	
	3	47 (23.7)	161 (21.9)	
	4	23 (11.6)	59 (8.0)	
	5	44 (22.2)	133 (18.1)	
Bottled water[$]	1	106 (52.5)	448 (58.0)	0.041
	2	21 (10.4)	45 (11.0)	
	3	42 (20.8)	99 (12.8)	
	4	33 (16.3)	140 (18.1)	
Pasteurised milk	1	50 (24.9)	140 (19.0)	0.013
	2	52 (25.9)	134 (18.2)	
	3	31 (15.4)	160 (21.7)	
	4	37 (18.4)	172 (23.3)	
	5	31 (15.4)	131 (17.8)	
Dairy products	1	38 (19.4)	145 (20.2)	0.067
	2	51 (26.0)	132 (18.4)	
	3	38 (19.4)	145 (20.2)	
	4	28 (14.3)	155 (21.6)	
	5	41 (20.9)	142 (19.8)	
Meat	1	30 (14.6)	173 (22.2)	0.042
	2	46 (22.4)	177 (22.8)	
	3	67 (32.7)	252 (32.4)	
	4	19 (9.3)	68 (8.7)	
	5	43 (21.0)	108 (13.9)	
Fish	1	54 (26.5)	145 (19.0)	0.041
	2	31 (15.2)	160 (21.0)	
	3	31 (15.2)	170 (22.3)	
	4	41 (20.1)	145 (19.0)	
	5	47 (23.0)	143 (18.7)	
Fruit	1	64 (31.2)	129 (17.0)	0.0001
	2	42 (20.5)	151 (19.9)	
	3	37 (18.1)	156 (20.5)	
	4	33 (16.1)	160 (21.1)	
	5	29 (14.2)	164 (21.6)	

* P value using chi square or exact test.
$ Quintiles automatically generated with the "xtile" command in Stata. Four categories were generated with the first group including observations up to the 40th percentile because 57 % of individuals had near zero intake.

Table 4 Odds ratios (95 % CI) comparing cases (n=218) and controls (n=812)

Variable		Univariable OR (95 % CI)	Adjusted OR (95 % CI)*	Multivariable OR (95 % CI)#
Age (yrs)		0.97 (0.96 – 0.98)	0.97 (0.96 – 0.98)	0.97 (0.95 - 0.98)[a]
Gender	Females	1.23 (0.91 – 1.67)	1.22 (0.89 – 1.67)	
Centre	Norwich	1.0	1.0	1.0[b]
	Leicester	1.99 (1.33 – 2.98)	1.99 (1.32 – 3.02)	2.09 (1.23 – 3.54)
	Sheffield	1.32 (0.81 – 2.15)	1.19 (0.72 – 1.97)	1.77 (0.97 – 3.21)
	Liverpool	1.81 (1.14 – 2.86)	1.99 (1.24 – 3.18)	2.40 (1.31 – 4.41)
	Bristol	1.90 (1.17 – 3.09)	1.73 (1.05 – 2.85)	2.12 (1.15 – 3.91)
OCP	Yes	1.28 (0.86 – 1.92)	1.03 (0.99 – 1.07)	
Ever smoke	Yes	1.82 (1.34 – 2.48)		
Smoke in YOI	Yes	1.34 (1.21 – 1.49)	1.27 (1.14 – 1.41)	1.31 (1.12 - 1.53)[c]
Smoking frequency	None	1.0		
	1 – 9	1.89 (1.14 – 3.09)		
	10 – 19	2.88 (1.86 – 4.44)		
	20 +	2.36 (1.46 – 3.82)		
Appendectomy	Yes	1.06 (0.67 – 1.68)	1.29 (0.80 – 2.08)	
Family history	Yes	4.37 (2.63 – 7.25)	4.20 (2.47 – 7.17)	7.13 (3.37 - 15.08)[a]
Water supply	Private	1.0	1.0	
	Public	1.35 (0.39 – 4.72)	1.6 (0.44 – 5.74)	
Water filter	Yes	0.45 (0.27 – 0.76)	0.54 (0.32 – 0.92)	
Wash dishes	Machine	1.0	1.0	
	Hand	1.21 (0.84 – 1.74)	1.27 (0.87 – 1.84)	
Holiday abroad	Yes	0.76 (0.56 – 1.04)	0.56 (0.40 – 0.79)	0.52 (0.32 - 0.84)[d]
Farm holiday	Yes	1.04 (0.47 – 2.31)	0.93 (0.61 – 1.42)	
Length farm holiday	None	1.0		

	< 1 week	1.87 (0.63 – 5.54)		
	1 week or more	0.86 (0.24 – 3.06)		
Farm visit	Yes	1.02 (0.68 – 1.54)		
Frequency farm visit	None	1.0	1.0	
(per year)	1	1.67 (0.71 – 3.89)	0.67 (0.19 – 2.42)	
	2	0.39 (0.09 – 1.71)	0.60 (0.07 – 5.09)	
	2 – 364	1.11 (0.64 – 1.92)	8.06 (0.68 – 95.30)	
	Over 365	0.70 (0.20 – 2.44)	-	
Contact farm animals	Yes	1.28 (0.74 – 2.21)	1.03 (0.99 – 1.07)	
Boiled water (ml) Ψ		0.95 (0.85 – 1.06)	0.99 (0.88 – 1.11)	
Unboiled water (ml) Ψ		1.17 (1.05 - 1.31)	1.01 (0.90 – 1.15)	
Bottled water (ml) Ψ		1.06 (0.96 – 1.16)	0.99 (0.89 – 1.09)	
Pasteurised milk (ml) Ψ		0.86 (0.77 – 0.96)	0.94 (0.83 – 1.05)	0.82 (0.69 – 0.97)b
Fruit (g) Ψ		0.78 (0.70 – 0.87)	0.78 (0.70 – 0.88)	0.78 (0.67 - 0.92)d
Dairy products (g) Ψ		0.95 (0.85 – 1.06)	0.96 (0.85 – 1.07)	
Meat (g) Ψ		1.19 (1.06 – 1.34)	1.28 (1.14 – 1.45)	1.40 (1.17 - 1.67)a
Fish (g) Ψ		1.0 (0.90 – 1.12)	1.01 (0.91 – 1.13)	
Cod liver oil (g) Ψ		0.82 (0.73 – 0.92)	0.99 (0.98 – 1.01)	

a p=0.0001; b p=0.01; c p=0.001; d p=0.002
OCP, Oral contraceptive pill; YOI, Year of interest
* Controlled for age and centre
# Significant variables following stepwise multivariable logistic regression with backward elimination
Ψ Frequency of intake x quantity over a 30 day period

A comparison of the unadjusted and adjusted ORs between cases and controls for the questionnaire data are shown in table 4. All significant variables were adjusted for age and centre and presented as adjusted variables. In the multivariable logistic regression analysis a stepwise procedure was used to sequentially eliminate variables with p values greater than 0.2.

The consumption of unboiled water was significantly associated with CD (OR 1.17, 95 % CI 1.05 – 1.31) in the univariable model. In contrast, filtered water consumption was associated with a significant 55 % (95 % CI 0.27 – 0.76) reduction in risk. Neither water supply type nor drinking boiled or bottled water were shown to be risk factors. None of the water related variables remained in the multivariable analysis.

Intake of pasteurised milk was shown to be significantly associated with the CD in the univariable (OR 0.86, 95 % CI 0.77 – 0.96) and multivariable model (OR 0.82, 95 % CI 0.69 – 0.97; p=0.01), although other dairy products were shown to have no effect (table 4).

Table 4 also shows that meat was significantly associated with the development of CD (OR 1.40, 95 % CI 1.17 – 1.67; p=0.0001). The meat category included both beef and canned meat products, although the majority refers to consumption of beef. Due to the positive association with meat, unadjusted and adjusted univariable analyses were conducted on the separate meat variables. In the unadjusted model, beef consumption was associated with a significant increase in risk (OR 1.18, 95 % CI 1.05 – 1.31, p=0.004) whereas canned meat intake was not (OR 1.11, 95 % CI 0.99 – 1.25, p=0.06). In the adjusted model taking into account centre and age, both consumption of beef (OR 1.21, 95 % CI 1.09 – 1.36, p=0.001) and canned meat (OR 1.19, 95 % CI 1.06 – 1.34, p=0.004) were significantly associated with an increased risk of CD. Although unadjusted and adjusted analyses have been presented, these changes have not been included in

the final model due to a high level of collinearity between the two terms. Furthermore, these amendments are post hoc and were not part of the original study design.

In the multivariable model controlling for only the effects of age and centre (table 4), smoking in the year of interest was associated with a 1.27 fold (95 % CI 1.14 – 1.41) increased risk of CD and family history was associated with a four fold increase in risk (95 % CI 2.47 – 7.17). In the fully adjusted model, potential confounding factors which were also shown to be a significant risk included smoking, particularly during the year of interest (OR 1.31, 95 % CI 1.12 – 1.53, p=0.001), and family history (OR 7.13, 95 % CI 3.37 – 15.08, p=0.0001). However, family holidays outside the UK (OR 0.52, 95 % CI 0.32 – 0.84, p=0.002) and fruit (OR 0.78, 95 % CI 0.67 – 0.92, p=0.002) consumption were associated with a reduced risk (table 4).

There was no evidence of interaction between age or centre and either consumption of pasteurised milk or meat intake (LR test p>0.06).

Drinking water-related exposure variables

Water and animal density data derived from the GIS comparing exposure between cases and controls during the year of interest and three (2000), five (1998) and 10 (1993) years previously are shown in tables 5a to 5d, respectively. For each year a multivariable model using stepwise logistic regression was also employed, however due to significant collinearity between terms, no variables remained in the model with the p value set at 0.2. Therefore, a multivariable model was fitted for each time period controlling for the effect of recognised risk factors (smoking, family history), confounding (age) and other factors (centre). In table 5d, data from both the home and work addresses in 1993 were available and have therefore been amalgamated to provide a more complete picture of possible risk.

The average distance to the WTWs, as a measure of biofilm potential, was consistently longer for the cases than the controls during all years recorded (tables 5a – 5d). The water treatment variables were similar between the cases and controls during all four time periods studied; they were also comparable over time (tables 5a – 5d). However, the cases received a greater amount of surface water compared with controls (tables 5a – 5d).

The number of cattle as a measure of ground water animal density were higher for the controls than the cases in the year of interest (table 5a), 2000 (table 5b) and 1998 (table 5c), but not 1993 (table 5d). Similarly, pig density was higher for the control population compared with cases. In contrast, the number of sheep were similar for the case population compared with the controls for all years recorded (tables 5a – 5d). In comparison, the animal density for sheep and pigs for the surface water source was greater for the controls than the cases during the year of interest, 2000 and 1998 (tables 5a and 5c, respectively). However, during 1993 the opposite trend was observed where the animal density was higher in the case population (tables 5b and 5d), except for sheep density. Cattle density for all years was higher for the cases compared with the control population.

The geological class indicates greater levels of permeability for the control population, similarly, the average borehole depth was greater for the controls compared with the cases for all years recorded (tables 5a – 5d). The county level reports of Johne's disease in cattle between 2000 and 2002 was also comparable between the cases and controls for all time periods (tables 5a – 5d) and did not vary significantly between 1993 and 2000 (tables 5b – 5d). None of the GIS derived exposure variables were significant in the univariable regression models.

Table 5a Water type, water treatment, animal density and ground protection data derived from the GIS comparing cases and controls during the year of interest. Values are shown as means (SD) and univariable ORs (95 % CI).

Variable	Variable subtype	Cases (n=170)	Controls (n=614)	Univariable OR (95 % CI)	Multivariable OR (95% CI)*
Sedimentation		0.03 (0.13)	0.03 (0.14)	0.90 (0.40 – 2.03)	1.20 (0.46 – 3.14)
Coagulation		0.04 (0.19)	0.07 (0.24)	0.63 (0.39 – 1.02)	0.71 (0.40 – 1.24)
Flocculation		0.26 (0.40)	0.27 (0.40)	0.88 (0.58 – 1.31)	0.89 (0.56 – 1.42)
Filtration		0.88 (0.26)	0.86 (0.27)	1.13 (0.76 – 1.69)	0.89 (0.57 – 1.41)
Distance to WTWs (km)		32.52 (21.72)	29.73 (21.06)	1.00 (0.999 – 1.000)	1.00 (0.999 – 1.000)
Water source~		1.54 (0.71)	1.48 (0.74)	1.15 (0.95 – 1.38)	1.08 (0.86 – 1.37)
Ground water animal density per 100,000 km^2	Cattle	0.46 (0.27)	0.51 (0.48)	1.03 (0.48 – 2.21)	0.98 (0.32 – 2.72)
	Sheep	0.65 (0.45)	0.65 (0.51)	1.08 (0.46 – 2.51)	1.58 (0.68 – 3.66)
	Pigs	0.98 (2.30)	1.40 (2.79)	0.99 (0.76 – 1.30)	0.82 (0.66 – 1.02)
Surface water animal density per 100,000 km^2	Cattle	1.08 (0.93)	1.00 (0.76)	0.82 (0.54 – 1.24)	1.02 (0.78 – 1.34)
	Sheep	3.44 (2.53)	3.54 (2.86)	0.96 (0.87 – 1.05)	1.00 (0.94 – 1.07)
	Pigs	0.46 (1.24)	0.65 (1.62)	0.89 (0.65 – 1.20)	1.08 (0.77 – 1.52)
Geological class		0.97 (0.87)	1.09 (1.07)	0.97 (0.84 – 1.12)	0.97 (0.82 – 1.13)
Drift material		0.03 (0.10)	0.05 (0.15)	0.78 (0.43 – 1.41)	0.80 (0.40 – 1.60)
Borehole depth (m)		1.6 (4.6)	2.8 (6.1)	0.98 (0.97 – 1.004)	0.98 (0.96 – 1.01)
Johne's disease in catchment		1.4 (1.1)	1.5 (1.6)	0.95 (0.88 – 1.23)	0.96 (0.86 – 1.06)

* Multivariable model controlling for the effect of age, centre, family history, smoking in the year of interest
~ score – ground - 0, mixed - 1, surface - 2
ORs for age, centre, family history and smoking are presented in table 4

Table 5b Water type, water treatment, animal density and ground protection data derived from the GIS comparing cases and controls during the year 2000. Values are shown as means (SD) and univariable ORs (95 % CI).

Variable	Variable subtype	Cases (n=164)	Controls (n=602)	Univariable OR (95 % CI)	Multivariable OR (95% CI)*
Sedimentation		0.03 (0.14)	0.03 (0.14)	0.98 (0.45 – 2.13)	1.36 (0.53 – 3.48)
Coagulation		0.05 (0.21)	0.08 (0.24)	0.76 (0.48 – 1.21)	0.90 (0.52 – 1.57)
Flocculation		0.26 (0.40)	0.27 (0.40)	0.87 (0.57 – 1.32)	0.81 (0.48 – 1.37)
Filtration		0.87 (0.26)	0.85 (0.28)	1.10 (0.74 – 1.64)	0.93 (0.58 – 1.50)
Distance to WTWs (km)		30.95 (20.74)	29.29 (20.94)	1.00 (0.999 – 1.000)	1.00 (0.999 – 1.000)
Water source~		1.56 (0.69)	1.48 (0.74)	1.14 (0.95 – 1.38)	1.12 (0.85 – 1.46)
Ground water animal density per 100,000 km^2	Cattle	0.45 (0.28)	0.52 (0.49)	1.07 (0.52 – 2.20)	0.91 (0.32 – 2.58)
	Sheep	0.62 (0.48)	0.65 (0.52)	0.80 (0.35 – 1.86)	1.00 (0.44 – 2.30)
	Pigs	0.98 (2.29)	1.46 (2.86)	1.04 (0.80 – 1.34)	0.77 (0.56 – 1.06)
Surface water animal density per 100,000 km^2	Cattle	1.05 (0.85)	1.00 (0.77)	1.01 (0.80 – 1.27)	1.04 (0.78 – 1.38)
	Sheep	3.37 (2.56)	3.56 (2.86)	1.00 (0.94 – 1.06)	0.99 (0.95 – 1.03)
	Pigs	0.50 (1.35)	0.65 (1.61)	1.05 (0.79 – 1.40)	1.03 (0.79 – 1.34)
Geological class		0.93 (0.87)	1.12 (1.10)	0.91 (0.78 – 1.06)	0.92 (0.77 – 1.11)
Drift material		0.03 (0.10)	0.06 (0.15)	0.73 (0.40 – 1.34)	0.83 (0.41 – 1.65)
Borehole depth (m)		1.6 (4.6)	2.9 (6.1)	0.99 (0.97 – 1.01)	0.99 (0.96 – 1.01)
Johne's disease in catchment		1.4 (1.1)	1.5 (1.5)	0.95 (0.88 – 1.03)	0.97 (0.87 – 1.08)

* Multivariable model controlling for the effect of age, centre, family history, smoking in the year of interest
~ score - ground - 0, mixed - 1, surface - 2
ORs for age, centre, family history and smoking are presented in table 4

Table 5c Water type, water treatment, animal density and ground protection data derived from the GIS comparing cases and controls during the year 1998. Values are shown as means (SD) and univariable ORs (95 % CI).

Variable	Variable subtype	Cases (n=151)	Controls (n=585)	Univariable OR (95 % CI)	Multivariable OR (95% CI)*
Sedimentation		0.03 (0.12)	0.03 (0.14)	0.73 (0.31 – 1.70)	1.09 (0.41 – 2.95)
Coagulation		0.05 (0.21)	0.07 (0.23)	0.78 (0.48 – 1.26)	0.98 (0.55 – 1.74)
Flocculation		0.26 (0.40)	0.27 (0.41)	0.89 (0.58 – 1.36)	0.89 (0.52 – 1.52)
Filtration		0.86 (0.27)	0.85 (0.28)	1.04 (0.69 – 1.57)	0.93 (0.58 – 1.51)
Distance to WTWs (km)		32.22 (20.06)	29.14 (20.93)	1.000 (0.999 – 1.000)	1.000 (0.999 – 1.000)
Water source~		1.57 (0.67)	1.47 (0.75)	1.15 (0.94 – 1.39)	1.09 (0.83 – 1.44)
Ground water animal density per 100,000 km^2	Cattle	0.46 (0.28)	0.49 (0.40)	1.06 (0.43 – 2.60)	1.02 (0.35 – 3.00)
	Sheep	0.64 (0.48)	0.63 (0.45)	1.01 (0.43 – 2.36)	1.20 (0.48 – 3.04)
	Pigs	0.98 (2.24)	1.35 (2.72)	0.97 (0.73 – 1.29)	0.84 (0.60 – 1.19)
Surface water animal density per 100,000 km^2	Cattle	1.04 (0.82)	1.00 (0.76)	0.92 (0.70 – 1.20)	0.96 (0.71 – 1.29)
	Sheep	3.44 (2.47)	3.55 (2.82)	0.98 (0.92 – 1.05)	0.98 (0.93 – 1.02)
	Pigs	0.48 (1.30)	0.62 (1.58)	0.94 (0.66 – 1.34)	0.95 (0.70 – 1.29)
Geological class		0.92 (0.89)	1.09 (1.07)	0.96 (0.81 – 1.13)	0.94 (0.78 – 1.13)
Drift material		0.03 (0.09)	0.05 (0.14)	0.69 (0.36 – 1.33)	0.82 (0.39 – 1.73)
Borehole depth (m)		1.5 (4.5)	2.8 (6.0)	0.99 (0.97 – 1.001)	0.99 (0.97 – 1.01)
Johne's disease in catchment		1.4 (1.0)	1.5 (1.5)	0.95 (0.88 – 1.03)	0.94 (0.97 – 1.01)

* Multivariable model controlling for the effect of age, centre, family history, smoking in the year of interest
~ score - ground - 0, mixed - 1, surface - 2
ORs for age, centre, family history and smoking are presented in table 4

Table 5d Water type, water treatment, animal density and ground protection data derived from the GIS comparing cases and controls during the year 1993. Values are shown as means (SD) and univariable ORs (95 % CI).

Variable	Variable subtype	Cases (n=167)	Controls (n=584)	Univariable OR (95 % CI)	Multivariable OR (95% CI)*
Sedimentation		0.02 (0.11)	0.03 (0.14)	0.59 (0.26 – 1.36)	0.84 (0.31 – 2.28)
Coagulation		0.06 (0.21)	0.08 (0.23)	0.71 (0.47 – 1.09)	0.96 (0.58 – 1.58)
Flocculation		0.25 (0.39)	0.28 (0.39)	0.88 (0.58 – 1.30)	0.73 (0.44 – 1.20)
Filtration		0.86 (0.26)	0.86 (0.26)	0.87 (0.61 – 1.25)	0.78 (0.51 – 1.18)
Distance to WTWs (km)		31.30 (19.73)	28.77 (20.69)	1.000 (0.999 – 1.000)	0.999 (0.999 – 1.000)
Water source~		1.57 (0.66)	1.51 (0.70)	1.10 (0.90 – 1.33)	0.995 (0.769 – 1.286)
Ground water animal density per 100,000 km^2	Cattle	0.49 (0.49)	0.46 (0.39)	1.61 (0.66 – 3.92)	1.75 (0.62 – 4.97)
	Sheep	0.67 (0.65)	0.61 (0.45)	1.36 (0.71 – 2.62)	1.74 (0.79 – 3.87)
	Pigs	0.82 (1.95)	1.14 (2.44)	1.01 (0.79 – 1.28)	0.88 (0.65 – 1.18)
Surface water animal density per 100,000 km^2	Cattle	1.11 (0.82)	1.02 (0.73)	0.80 (0.55 – 1.16)	0.89 (0.66 – 1.20)
	Sheep	3.71 (2.72)	3.68 (2.62)	0.94 (0.86 – 1.03)	0.96 (0.90 – 1.02)
	Pigs	0.63 (1.58)	0.74 (1.76)	0.82 (0.57 – 1.17)	0.78 (0.56 – 1.10)
Geological class		0.96 (0.85)	1.00 (0.97)	0.99 (0.87 – 1.11)	0.96 (0.85 – 1.08)
Drift material	Yes	0.02 (0.08)	0.04 (0.12)	0.55 (0.27 – 1.11)	0.65 (0.29 – 1.45)
Borehole depth (m)		1.5 (4.4)	2.4 (5.4)	0.99 (0.97 – 1.01)	0.99 (0.97 – 1.01)
Johne's disease in catchment		1.5 (1.1)	1.6 (2.0)	0.96 (0.91 – 1.01)	0.99 (0.94 – 1.03)

* Multivariable model controlling for the effect of age, centre, family history, smoking in the year of interest
~ score - ground - 0, mixed - 1, surface - 2
ORs for age, centre, family history and smoking are presented in table 4

Sub group analysis

A comparison of exposures for the cases with the type of CD categorised as stricturing, penetrating and non-stricturing non-penetrating are displayed in tables 6a to 6c, respectively. A description of each category can be found in appendix 3. The type of water supply and consumption of drinking water, whether filtered, boiled or bottled was not associated with a significantly increased risk for any of the three disease types (tables 6a – 6c). However, a 1.5 fold increased risk was shown for unboiled water intake in cases categorised with penetrating CD (table 6b), although this effect was not reflected in the multivariable model.

Pasteurised milk intake was associated with a 30 % reduction in risk (95 % CI 0.52 – 0.94) in cases with penetrating CD (table 6b), but this effect was eliminated in the multivariable model. A similar decreased risk was also shown in the stricturing group (table 6a), although this reduction was not significant (OR 0.82, 95 % CI 0.68 – 1.00). Dairy products were not a significant risk factor for any of the disease types (tables 6a – 6c).

In cases with non-stricturing non-penetrating disease (table 6c), eating meat was associated with a 1.2 fold (95 % CI 1.04 – 1.40) increased risk, which rose to about 1.5 fold (95 % CI 1.19 – 1.85, p=0.001) in the multivariable model. However, a similar trend was not observed for either the stricturing or penetrating disease groups (tables 6a and 6b). Similarly, although fish intake was not significant in the univariable model, it was associated with a 22 % reduction in risk (95 % CI 0.63 – 0.95, p=0.014) in the multivariable analysis for cases categorised with non-stricturing non-penetrating disease (table 6c). However, a similar trend was not observed in either the stricturing or penetrating groups (tables 6a and 6b). Fruit intake was also shown to decrease risk of disease. In the multivariable model, cases with penetrating and non-stricturing non-penetrating CD demonstrated a 35 % (95 % CI 0.44 – 0.96, p=0.029; table 5b) and 22 % (95 % CI 0.63 – 0.95,

Table 6a Comparison of types of exposure for cases with stricturing CD. Values are shown as total (%) of cases with stricturing disease and ORs (95 % CI), unless otherwise stated

Variable		Cases (n=63)	Univariable OR (95 % CI)	Multivariable OR (95 % CI)
Age (yrs)	Mean (SD)	42 (19.1)	0.97 (0.96 – 0.99)	0.97 (0.94 – 0.99)[a]
Gender	Females	34/63 (54.0)	1.38 (0.83 – 2.30)	
Centre	Norwich	28/63 (44.4)	1.0	
	Leicester	13/63 (20.6)	1.14 (0.58 – 2.27)	
	Sheffield	8/63 (12.7)	0.92 (0.41 – 2.07)	
	Liverpool	8/63 (12.7)	0.98 (0.43 – 2.21)	
	Bristol	6/63 (9.5)	0.89 (0.36 – 2.23)	
OCP	Yes	20/33 (60.6)	1.05 (0.99 – 1.12)	
Smoke in YOI	Yes	31/61 (50.8)	1.44 (1.22 – 1.72)	1.48 (1.16 – 1.89)[a]
Appendectomy	Yes	10/62 (16.1)	1.39 (0.68 – 2.82)	
Family history	Yes	9/57 (15.8)	3.55 (1.61 – 7.84)	5.02 (1.66 – 15.14)[b]
Water supply	Private	1/61 (1.6)	1.0	
	Public	60/61 (98.4)	1.16 (0.15 – 8.95)	
Water filter	Yes	4/56 (7.1)	0.36 (0.13 – 1.00)	
Wash dishes	Machine	13/62 (21.0)	1.0	
	Hand	49/62 (79.0)	1.26 (0.67 – 2.37)	
Holiday abroad	Yes	35/62 (56.5)	0.69 (0.41 – 1.16)	0.40 (0.19 – 0.85)[c]
Farm holiday	Yes	1/63 (1.6)	1.46 (0.76 – 2.79)	
Contact farm animal	Yes	7/19 (36.8)	1.05 (0.99 – 1.12)	
Boiled water (ml)	Mean (SD) [Ψ]	26664 (16343)	0.95 (0.79 – 1.15)	
Unboiled water (ml)	Mean (SD) [Ψ]	12261 (10492)	1.09 (0.90 – 1.31)	
Bottled water (ml)	Mean (SD) [Ψ]	475 (328)	1.02 (0.87 – 1.19)	
Pasteurised milk (ml)	Mean (SD) [Ψ]	21564 (18853)	0.82 (0.68 – 1.00)	
Fruit (g)	Mean (SD) [Ψ]	6601 (5287)	0.78 (0.65 – 0.95)	
Dairy products (g)	Mean (SD) [Ψ]	3200 (2069)	0.87 (0.72 – 1.04)	
Meat (g)	Mean (SD) [Ψ]	1823 (3303)	1.20 (0.98 – 1.45)	
Fish (g)	Mean (SD) [Ψ]	1661 (1310)	1.06 (0.88 – 1.29)	
Cod liver oil (g)	Mean (SD) [Ψ]	1.6 (1.2)	0.99 (0.97 – 1.01)	

a p=0.002; b p=0.004; c p=0.018
OCP, Oral contraceptive pill; YOI, Year of interest
[Ψ] Frequency of intake x quantity over a 30 day period

Table 6b Comparison of types of exposure for cases with penetrating CD. Values shown as total (%) of cases with penetrating disease and ORs (95 % CI), unless otherwise stated

Variable		Cases (n=29)	Univariable OR (95 % CI)	Multivariable OR (95 % CI)
Age (yrs)	Mean (SD)	38 (19.4)	0.96 (0.94 – 0.98)	0.96 (0.93 – 0.99)[a]
Gender	Females	17/29 (58.6)	1.10 (0.53 – 2.33)	
Centre	Norwich	7/29 (24.1)	1.0	1.40 (1.01 – 1.93)[b]
	Leicester	7/29 (24.1)	2.46 (0.85 – 7.13)	
	Sheffield	4/29 (13.8)	1.83 (0.53 – 6.37)	
	Liverpool	3/29 (10.3)	1.47 (0.37 – 5.77)	
	Bristol	8/29 (27.6)	4.77 (1.68 – 13.52)	
OCP	Yes	10/17 (58.8)	1.02 (0.93 – 1.12)	
Smoke in YOI	Yes	11/29 (37.9)	1.26 (0.98 – 1.61)	
Appendectomy	Yes	5/29 (17.2)	1.47 (0.55 – 3.94)	
Family history	Yes	6/27 (22.2)	5.27 (2.00 – 13.88)	9.75 (2.50 – 38.00)[c]
Water supply	Private	2/29 (6.9)	1.0	
	Public	27/29 (93.1)	0.27 (0.06 – 1.22)	
Water filter	Yes	1/28 (3.6)	0.17 (0.02 – 1.25)	
Wash dishes	Machine	4/29 (13.8)	1.0	
	Hand	25/29 (86.2)	2.13 (0.73 – 6.18)	
Holiday abroad	Yes	15/28 (53.6)	0.64 (0.30 – 1.34)	0.34 (0.12 – 0.93)[d]
Farm holiday	Yes	2/28 (7.1)	2.17 (0.97 – 4.88)	
Contact farm animal	Yes	5/14 (35.7)	0.96 (0.88 – 1.04)	
Boiled water (ml)	Mean (SD) [Ψ]	21522 (19583)	0.86 (0.65 – 1.13)	
Unboiled water (ml)	Mean (SD) [Ψ]	19053 (12577)	1.48 (1.13 – 1.94)	
Bottled water (ml)	Mean (SD) [Ψ]	530 (324)	1.12 (0.89 – 1.41)	
Pasteurised milk (ml)	Mean (SD) [Ψ]	18126 (14736)	0.70 (0.52 – 0.94)	
Fruit (g)	Mean (SD) [Ψ]	5530 (4474)	0.63 (0.47 – 0.86)	0.65 (0.44 – 0.96)[e]
Dairy products (g)	Mean (SD) [Ψ]	3765 (1934)	1.09 (0.82 – 1.44)	
Meat (g)	Mean (SD) [Ψ]	1607 (2873)	1.13 (0.85 – 1.50)	
Fish (g)	Mean (SD) [Ψ]	1333 (904)	0.83 (0.63 – 1.11)	
Cod liver oil (g)	Mean (SD) [Ψ]	1.1 (0.6)	0.93 (0.86 – 1.00)	

a p=0.013; b p=0.042; c p=0.001; d p= 0.037; e p=0.029
OCP, Oral contraceptive pill; YOI, Year of interest
[Ψ] Frequency of intake x quantity over a 30 day period

Table 6c Comparison of types of exposure for cases with non-stricturing non-penetrating CD. Values shown as total (%) of cases with non-stricturing non-penetrating disease and ORs (95 % CI), unless otherwise stated.

Variable		Cases (n=125)	Univariable OR (95 % CI)	Multivariable OR (95 % CI)
Age (yrs)	Mean (SD)	42 (19.4)	0.98 (0.96 – 0.99)	0.96 (0.95 – 0.98)[a]
Gender	Females	73/125 (58.4)	1.17 (0.80 – 1.71)	
Centre	Norwich	33/125 (26.4)	1.0	
	Leicester	35/125 (28.0)	2.61 (1.56 – 4.36)	
	Sheffield	16/125 (12.8)	1.55 (0.82 – 2.93)	
	Liverpool	24/125 (19.2)	2.49 (1.41 – 4.39)	
	Bristol	17/125 (13.6)	2.15 (1.14 – 4.04)	4.11 (1.11 – 15.31)[b]
OCP	Yes	45/73 (61.6)	1.03 (0.98 – 1.08)	
Smoke in YOI	Yes	52/122 (42.6)	1.32 (1.16 – 1.50)	1.30 (1.07 – 1.57)[c]
Appendectomy	Yes	12/122 (9.8)	0.80 (0.42 – 1.50)	
Family history	Yes	20/105 (19.1)	4.50 (2.47 – 8.19)	8.5 (3.70 – 19.50)[a]
Water supply	Private	0/125 (0)	- *	
	Public	125/125 (100)	- *	
Water filter	Yes	13/120 (10.8)	0.57 (0.31 – 1.03)	
Wash dishes	Machine	29/124 (23.4)	1.0	
	Hand	95/124 (76.6)	1.09 (0.70 – 1.69)	
Holiday abroad	Yes	75/121 (62.0)	0.86 (0.58 – 1.27)	
Farm holiday	Yes	5/121 (4.1)	0.70 (0.40 – 1.25)	
Contact farm animal	Yes	11/40 (27.5)	1.03 (0.99 – 1.08)	
Boiled water (ml)	Mean (SD) [Ψ]	27146 (17341)	0.97 (0.84 – 1.12)	
Unboiled water (ml)	Mean (SD) [Ψ]	14735 (15136)	1.12 (0.98 – 1.29)	
Bottled water (ml)	Mean (SD) [Ψ]	503 (332)	1.07 (0.95 – 1.20)	
Pasteurised milk (ml)	Mean (SD) [Ψ]	24330 (15658)	0.92 (0.80 – 1.06)	
Fruit (g)	Mean (SD) [Ψ]	6941 (6627)	0.80 (0.69 – 0.92)	0.78 (0.63 – 0.95)[d]
Dairy products (g)	Mean (SD) [Ψ]	3450 (2209)	0.94 (0.82 – 1.09)	
Meat (g)	Mean (SD) [Ψ]	2453 (4369)	1.21 (1.04 – 1.40)	1.48 (1.19 – 1.85)[e]
Fish (g)	Mean (SD) [Ψ]	1629 (1219)	1.00 (0.87 – 1.14)	0.78 (0.63 – 0.95)[f]
Cod liver oil (g)	Mean (SD) [Ψ]	1.6 (1.2)	0.99 (0.97 – 1.00)	

a p=0.000; b p=0.035; c p=0.008; d p=0.013; e p=0.001; f p=0.014
* ORs were not calculated for water supply since all cases received a public supply
OCP, Oral contraceptive pill; YOI, Year of interest
[Ψ] Frequency of intake x quantity over a 30 day period

p=0.013; table 6c) reduction in risk, respectively. Similarly, fruit intake was associated with a decreased risk of CD in cases with stricturing disease (OR 0.78, 95 % CI 0.65 – 0.95), however this effect was only observed in the univariable logistic regression model (table 6a). Disease risk was also predicted by age of onset for all disease types (tables 6a to 6c).

For both logistic regression models, family history was shown to have a significant increased risk in stricturing (OR 5.02, 95 % CI 1.66 – 15.14, p=0.004; table 6a), penetrating (OR 9.75, 95 % CI 2.50 – 38.00, p=0.001; table 6b) and non-stricturing non-penetrating (OR 8.5, 95 % CI 3.70 – 19.50, p=0.000; table 6c) CD. A similar trend was seen for smoking in the year of interest with a 1.5 fold and 1.3 fold increased risk of CD in cases categorised with stricturing (95 % CI 1.16 – 1.89, p=0.002; table 6a) and non-stricturing non-penetrating (95 % CI 1.07 – 1.57, p=0.008; table 6c) disease, respectively. However, in cases with penetrating CD, smoking was not shown to be a significant risk factor in either the univariable or multivariable analysis (table 6b). Holidays outside the UK was a significant predictor of a decrease in risk in cases with stricturing (OR 0.40, 95 % CI 0.19 – 0.85, p=0.018; table 6a) and penetrating (OR 0.34, 95 % CI 0.12 – 0.93, p=0.037; table 6b) CD for both logistic models, but not for those with non-stricturing non-penetrating disease (table 6c).

Tables 7a to 7d compare the types of exposure derived from the questionnaire data for cases with CD of the upper GI (7a), terminal ileum (7b), ileocolon (7c) and colon (7d). A description of each disease site can be found in appendix 3. Water supply and consumption of filtered, boiled or bottled water was not associated with a risk of CD for any of the disease sites (tables 7a – 7d). However, consuming unboiled water increased risk by 1.4 fold (95 % CI 1.12 – 1.85) in cases with CD of the upper GI (table 7a), although this effect was eliminated in the multivariable analysis.

Table 7a Comparison of types of exposure for cases with CD of the upper GI. Values shown as total (%) of cases with CD of the upper GI and ORs (95 % CI), unless otherwise stated.

Variable		Cases (n=27)	Univariable OR (95 % CI)	Multivariable OR (95 % CI)
Age (yrs)	Mean (SD)	36 (14.8)	0.95 (0.94 – 0.97)	0.93 (0.90 – 0.98)[a]
Gender	Females	11/27 (40.7)	1.86 (0.95 – 3.63)	
Centre	Norwich	6/27 (22.2)	1.0	
	Leicester	7/27 (25.9)	2.87 (0.95 – 8.68)	
	Sheffield	2/27 (7.4)	1.07 (0.21 – 5.37)	
	Liverpool	7/27 (25.9)	3.99 (1.31 – 12.13)	
	Bristol	5/27 (18.5)	3.48 (1.04 – 11.66)	
OCP	Yes	9/11 (81.8)	1.09 (1.01 – 1.19)	1.17 (1.01 – 1.36)[b]
Smoke in YOI	Yes	16/27 (59.3)	1.55 (1.24 – 1.94)	
Appendectomy	Yes	6/27 (22.2)	1.78 (0.76 – 4.17)	
Family history	Yes	3/21 (14.3)	4.03 (1.45 – 11.23)	5.72 (1.15 – 28.6)[c]
Water supply	Private	0/27 (0)	- *	
	Public	27/27 (100)	- *	
Water filter	Yes	1/24 (4.2)	0.49 (0.15 – 1.62)	
Wash dishes	Machine	4/27 (14.8)	1.0	
	Hand	23/27 (85.2)	1.64 (0.67 – 4.00)	
Holiday abroad	Yes	16/26 (61.5)	0.95 (0.46 – 1.95)	
Farm holiday	Yes	3/27 (11.1)	3.14 (1.52 – 6.47)	3.55 (1.17 – 10.79)[d]
Contact farm animal	Yes	7/10 (70.0)	0.99 (0.91 – 1.07)	
Boiled water (ml)	Mean (SD) [Ψ]	29140 (14817)	1.05 (0.82 – 1.34)	
Unboiled water (ml)	Mean (SD) [Ψ]	20178 (13169)	1.44 (1.12 – 1.85)	
Bottled water (ml)	Mean (SD) [Ψ]	556 (361)	1.17 (0.92 – 1.49)	
Pasteurised milk (ml)	Mean (SD) [Ψ]	23903 (18474)	0.88 (0.68 – 1.14)	
Fruit (g)	Mean (SD) [Ψ]	7678 (6379)	0.83 (0.65 – 1.08)	
Dairy products (g)	Mean (SD) [Ψ]	3178 (2283)	0.89 (0.69 – 1.14)	
Meat (g)	Mean (SD) [Ψ]	1150 (1491)	1.12 (0.86 – 1.45)	
Fish (g)	Mean (SD) [Ψ]	1488 (1155)	1.00 (0.78 – 1.28)	
Cod liver oil (g)	Mean (SD) [Ψ]	1.9 (1.4)	1.00 (0.98 – 1.02)	

a p=0.002; b p=0.034; c p=0.033; d p=0.025;
* ORs were not calculated for water supply since all cases received a public supply
OCP, Oral contraceptive pill; YOI, Year of interest
[Ψ] Frequency of intake x quantity over a 30 day period

Table 7b Comparison of types of exposure for cases with CD of the terminal ileum. Values shown as total (%) of cases with CD of the terminal ileum and ORs (95 % CI), unless otherwise stated.

Variable		Cases (n=14)	Univariable OR (95 % CI)	Multivariable OR (95 % CI)
Age (yrs)	Mean (SD)	40 (19.0)	0.96 (0.94 – 0.99)	
Gender	Females	8/14 (57.1)	1.07 (0.46 – 2.50)	
Centre	Norwich	6/14 (42.9)	1.0	
	Leicester	2/14 (14.3)	0.82 (0.16 – 41.11)	
	Sheffield	0/14 (0)	-	
	Liverpool	3/14 (21.4)	1.71 (0.42 – 6.95)	
	Bristol	3/14 (21.4)	2.09 (0.51 – 8.51)	
OCP	Yes	6/7 (85.7)	1.03 (0.93 – 1.15)	
Smoke in YOI	Yes	7/14 (50.0)	1.40 (1.06 – 1.85)	
Appendectomy	Yes	1/13 (7.7)	0.74 (0.17 – 3.20)	
Family history	Yes	1/12 (8.3)	3.41 (0.95 – 12.22)	20.20 (2.47 – 165.34)[a]
Water supply	Private	0/13 (0)	- *	
	Public	13/13 (100)	- *	
Water filter	Yes	0/11 (0)	0.56 (0.13 – 2.42)	
Wash dishes	Machine	3/14 (21.4)	1.0	
	Hand	11/14 (78.6)	1.18 (0.43 – 3.22)	
Holiday abroad	Yes	8/14 (57.1)	0.91 (0.38 – 2.19)	
Farm holiday	Yes	0/14 (0)	1.51 (0.54 – 4.19)	
Contact farm animal	Yes	1/4 (25.0)	1.00 (0.90 – 1.10)	
Boiled water (ml)	Mean (SD) [Ψ]	25910 (23790)	1.10 (0.80 – 1.52)	
Unboiled water (ml)	Mean (SD) [Ψ]	14370 (10985)	1.26 (0.92 – 1.73)	
Bottled water (ml)	Mean (SD) [Ψ]	500 (335)	1.06 (0.76 – 1.49)	
Pasteurised milk (ml)	Mean (SD) [Ψ]	26529 (17282)	0.93 (0.68 – 1.29)	
Fruit (g)	Mean (SD) [Ψ]	7995 (9763)	0.69 (0.50 – 0.96)	
Dairy products (g)	Mean (SD) [Ψ]	2317 (749)	0.81 (0.59 – 1.11)	
Meat (g)	Mean (SD) [Ψ]	1874 (4096)	1.05 (0.76 – 1.45)	
Fish (g)	Mean (SD) [Ψ]	1660 (1108)	1.11 (0.82 – 1.51)	
Cod liver oil (g)	Mean (SD) [Ψ]	1.5 (1.2)	0.99 (0.96 – 1.02)	

a p=0.005
* ORs were not calculated for water supply since all cases received a public supply
OCP, Oral contraceptive pill; YOI, Year of interest
[Ψ] Frequency of intake x quantity over a 30 day period

Table 7c Comparison of types of exposure for cases with CD of the ileocolon. Values shown as total (%) of cases with CD of the ileocolon and ORs (95 % CI), unless otherwise stated.

Variable		Cases (n=95)	Univariable OR (95 % CI)	Multivariable OR (95 % CI)
Age (yrs)	Mean (SD)	41 (18.8)	0.97 (0.96 – 0.98)	0.96 (0.95 – 0.98)[a]
Gender	Females	51/95 (53.7)	1.37 (0.91 – 2.07)	
Centre	Norwich	31/95 (32.6)	1.0	
	Leicester	23/95 (24.2)	1.82 (1.03 – 3.23)	
	Sheffield	19/95 (20.0)	1.96 (1.07 – 3.61)	3.51 (1.41 – 8.77)[f]
	Liverpool	14/95 (14.7)	1.54 (0.79 – 3.01)	
	Bristol	8/95 (8.4)	1.08 (0.48 – 2.43)	
OCP	Yes	29/51 (56.9)	1.04 (0.99 – 1.10)	
Smoke in YOI	Yes	41/93 (44.1)	1.34 (1.16 – 1.54)	1.33 (1.09 – 1.63)[b]
Appendectomy	Yes	13/95 (13.7)	1.14 (0.63 – 2.09)	
Family history	Yes	16/84 (19.1)	4.70 (2.52 – 8.77)	6.97 (2.98 – 16.34)[a]
Water supply	Private	2/95 (2.1)	1.0	
	Public	93/95 (97.9)	0.97 (0.22 – 4.31)	
Water filter	Yes	9/95 (9.5)	0.56 (0.29 – 1.08)	
Wash dishes	Machine	24/95 (25.3)	1.0	
	Hand	71/95 (74.7)	0.98 (0.61 – 1.58)	
Holiday abroad	Yes	53/93 (57.0)	0.73 (0.48 – 1.11)	0.36 (0.19 – 0.66)[c]
Farm holiday	Yes	3/93 (3.2)	0.94 (0.53 – 1.66)	
Contact farm animal	Yes	7/33 (21.2)	1.02 (0.97 – 1.07)	
Boiled water (ml)	Mean (SD) [Ψ]	22827 (15744)	0.88 (0.76 – 1.03)	
Unboiled water (ml)	Mean (SD) [Ψ]	13241 (13359)	1.10 (0.94 – 1.28)	
Bottled water (ml)	Mean (SD) [Ψ]	494 (323)	1.05 (0.92 – 1.20)	
Pasteurised milk (ml)	Mean (SD) [Ψ]	20586 (16795)	0.77 (0.66 – 0.91)	
Fruit (g)	Mean (SD) [Ψ]	6652 (6261)	0.76 (0.66 – 0.89)	0.78 (0.63 – 0.96)[d]
Dairy products (g)	Mean (SD) [Ψ]	3392 (2323)	0.92 (0.79 – 1.07)	
Meat (g)	Mean (SD) [Ψ]	2119 (3819)	1.33 (1.13 – 1.56)	1.58 (1.26 – 1.98)[a]
Fish (g)	Mean (SD) [Ψ]	1638 (1356)	1.02 (0.88 – 1.18)	0.82 (0.66 – 1.01)[e]
Cod liver oil (g)	Mean (SD) [Ψ]	1.3 (1.0)	0.97 (0.95 – 0.99)	

a p=0.000; b p=0.006; c p=0.001; d p=0.017; e p=0.056, f p=0.007
OCP, Oral contraceptive pill; YOI, Year of interest
[Ψ] Frequency of intake x quantity over a 30 day period

Table 7d Comparison of types of exposure for cases with CD of the colon. Values shown as total (%) of cases with CD of the colon and ORs (95 % CI), unless otherwise stated.

Variable		Cases (n=73)	Univariable OR (95 % CI)	Multivariable OR (95 % CI)
Age (yrs)	Mean (SD)	46 (21.0)	0.98 (0.97 – 0.99)	0.97 (0.95 – 0.98)[a]
Gender	Females	49/73 (67.1)	0.82 (0.50 – 1.32)	
Centre	Norwich	25/73 (34.3)	1.0	
	Leicester	18/73 (24.7)	1.77 (0.94 – 3.34)	
	Sheffield	6/73 (8.2)	0.77 (0.31 – 1.92)	
	Liverpool	10/73 (13.7)	1.37 (0.64 – 2.94)	
	Bristol	14/73 (19.2)	2.34 (1.17 – 4.69)	
OCP	Yes	29/49 (59.2)	0.98 (0.93 – 1.04)	
Smoke in YOI	Yes	27/70 (38.6)	1.24 (1.06 – 1.45)	
Appendectomy	Yes	6/70 (8.6)	0.72 (0.32 – 1.60)	
Family history	Yes	13/65 (20.0)	5.00 (2.57 – 9.74)	9.29 (3.73 – 23.16)[a]
Water supply	Private	1/72 (1.4)	1.0	
	Public	71/72 (98.6)	1.52 (0.20 – 11.69)	
Water filter	Yes	6/67 (9.0)	0.56 (0.26 – 1.20)	
Wash dishes	Machine	13/71 (18.3)	1.0	
	Hand	58/71 (81.7)	1.42 (0.79 – 2.55)	
Holiday abroad	Yes	43/71 (60.6)	0.85 (0.53 – 1.36)	
Farm holiday	Yes	2/70 (2.9)	0.73 (0.37 – 1.46)	
Contact farm animal	Yes	7/22 (31.8)	1.04 (0.98 – 1.10)	
Boiled water (ml)	Mean (SD) [Ψ]	28663 (18213)	1.06 (0.89 – 1.25)	
Unboiled water (ml)	Mean (SD) [Ψ]	14189 (14566)	1.13 (0.96 – 1.33)	
Bottled water (ml)	Mean (SD) [Ψ]	492 (332)	1.05 90.91 – 1.21)	
Pasteurised milk (ml)	Mean (SD) [Ψ]	24199 (15666)	0.97 (0.81 – 1.15)	
Fruit (g)	Mean (SD) [Ψ]	6146 (4579)	0.74 (0.62 – 0.88)	0.75 (0.59 – 0.94)[b]
Dairy products (g)	Mean (SD) [Ψ]	3621 (1915)	1.11 (0.94 – 1.32)	
Meat (g)	Mean (SD) [Ψ]	2450 (4309)	1.19 (0.99 – 1.41)	
Fish (g)	Mean (SD) [Ψ]	1548 (1069)	1.02 (0.86 – 1.20)	
Cod liver oil (g)	Mean (SD) [Ψ]	1.6 (1.2)	0.99 (0.97 – 1.01)	

a p=0.000; b p=0.014
OCP, Oral contraceptive pill; YOI, Year of interest
[Ψ] Frequency of intake x quantity over a 30 day period

In the univariable analysis, pasteurised milk intake was associated with a reduced risk (OR 0.77, 95 % CI 0.66 – 0.91) in cases identified with CD of the ileocolon (table 7c), although a similar observation was not seen for any of the other disease locations (tables 7a – 7b, 7d). For all cases no significant risk was associated with other dairy product intake.

Meat consumption was associated with a 1.58 fold (95 % CI 1.26 – 1.98, p=0.000) increased risk in cases with ileocolon CD (table 7c) after controlling for confounding in a multivariable model. However, a similar risk was not observed for cases with CD at any other site of the GIT. A significant reduction in risk of up to 26 % associated with fruit intake was noted in both the ileocolon (table 7c) and colon (table 7d) groups. This decreased risk was seen in both the univariable (ileocolon, OR 0.76, 95 % CI 0.66 – 0.89; colon, OR 0.74, 95 % CI 0.62 – 0.88) and multivariable (ileocolon, OR 0.78, 95 % CI 0.63 – 0.96, p=0.017; colon, OR 0.75, 95 % CI 0.59 – 0.94, p=0.014) logistic regression models. A similar effect was also observed in cases with CD of the terminal ileum (OR 0.69, 95 % CI 0.50 – 0.96), but this trend was subsequently eliminated from the multivariable analysis (table 7b). Fruit intake was not shown to have any significant effect in cases with CD of the upper GI (table 7a). A reduced risk in the ileocolon group (table 7c) was also associated with taking cod liver oil (OR 0.97, 95 % CI 0.95 – 0.99). However this effect did not remain significant in the multivariable analysis, and was not observed for cases with CD at any other site.

As with disease type, family history was associated with a significantly ($p<0.034$) increased risk of CD for all cases, regardless of disease site (tables 7a – 7d). This was most pronounced in the terminal ileum group (table 7b), where a 20 fold increase in risk was observed (95 % CI 2.47 – 165.34, p=0.005) in the multivariable analysis, although this effect was not significant in the univariable model. Smoking during the year of interest was also a significant predictor of risk, with ORs ranging between 1.24 (95 % CI 1.06 – 1.45) and 1.55 (95 % CI 1.24 – 1.94) in cases

with colon (table 7d) and upper GI (table 7a) CD, respectively. However in the multivariable model only smoking in the ileocolon group (table 7c) remained a significant risk (OR 1.33, 95 % CI 1.09 – 1.63, p=0.006). The use of the oral contraceptive pill was associated with an increased risk (OR 1.09, 95 % CI 1.01 – 1.19) which was also reflected in the multivariable model (OR 1.17, 95 % CI 1.01 – 1.36, p=0.034) for cases with CD of the upper GI (table 7a). A similar trend was not observed for any of the other disease sites (tables 7b – 7d).

In cases with CD of the upper GI (table 7a), taking farm holidays was associated with a three fold increase in the risk of developing the disease; an observation noted in both the univariable (OR 3.14, 95 % CI 1.52 – 6.47) and multivariable (OR 3.55, 95 % CI 1.17 – 10.79, p=0.025) models, but not for CD of the terminal ileum, ileocolon or colon. Furthermore, holidays outside the UK were associated with a 64 % reduction in risk (95 % CI 0.19 – 0.66, p=0.001) in the ileocolon group (table 7c). A similar observation was not seen in the univariable analysis, or for any of the other disease sites.

Table 8 shows a comparison of the exposure variables derived from the questionnaire data restricted to cases with onset of symptoms between June 2002 and June 2004 only, to assess the impact of recall bias. None of the water related or dairy product variables remained in the multivariable model. For dietary intake, eating meat was associated with an increased risk (OR 1.55, 95 % CI 1.26 – 1.91, p=0.001) in cases with onset of symptoms during the last two years of the study, whereas fruit was shown to decrease risk (OR 0.81, 95 % CI 0.67 – 0.97, p=0.02). However, family history was associated with a 6 fold increase in risk (95 % CI 2.75 – 15.40, p=0.001). Smoking during the year of interest was also shown to be a significant (p=0.005) predictor of increased risk in cases diagnosed during the last two years of the study (OR 1.29, 95 % CI 1.08 – 1.54). Furthermore, holidays outside the UK were associated with a 44 % reduction in risk (95 % CI 0.32 – 0.97, p=0.039) (table 8).

Table 8 Multivariate analysis of cases symptomatic within the last two years only (n=124)

	Odds Ratio (95 % CI)	P Value
Age	0.96 (0.94 - 0.97)	0.000
Gender	7.45 (0.72 - 77.07)	0.092
Family History	6.50 (2.75 - 15.40)	0.000
Centre	1.27 (1.07 - 1.50)	0.006
OCP	0.77 (0.57 - 1.03)	0.076
Smoke in YOI	1.29 (1.08 - 1.54)	0.005
Holiday Abroad	0.56 (0.32 - 0.97)	0.039
Pasteurised Milk	0.88 (0.72 - 1.06)	0.168
Fruit	0.81 (0.67 - 0.97)	0.02
Meat	1.55 (1.26 - 1.91)	0.000
Fish	0.86 (0.72 - 1.04)	0.115

OCP, Oral contraceptive pill; YOI, Year of interest
Multivariable regression with age, gender, family history, water supply, drinking filtered water, holiday abroad, farm holiday visit, contact with farm animals, quantity of boiled water, unboiled water, pasteurised milk, fish, meat, fruits, cod liver oil and dairy products consumed, smoking in the year of interest, use of oral contraceptives, appendectomy, dish washing practice and centre as explanatory variables.

Table 9 Awareness of the causes of CD among the study population (n=1030). Values shown as totals (%) of each group.

Causes	Cases (n=218)	Controls (n=812)
Bacteria	9/218 (4.1)	4/812 (0.5)
Dairy Products	10/218 (4.6)	10/812 (1.2)
Following Illness	5/218 (2.3)	2/812 (0.3)
Food Intolerance	12/218 (5.5)	15/812 (1.9)
Genetics	9/218 (4.1)	7/812 (0.9)
Medication	5/218 (2.3)	1/812 (0.1)
Smoking	10/218 (4.6)	6/812 (0.7)
Stress	26/218 (11.9)	10/812 (1.2)
Vaccinations	3/218 (1.4)	2/812 (0.3)
Viral Infections	2/218 (0.9)	4/812 (0.5)
Water	2/218 (0.9)	2/812 (0.3)
Unaware of cause	157/218 (72.0)	766/812 (94.3)

Other factors reported include: Animals – one case; Alcohol – two cases and one control; Irregular food intake, one case; ME one case

The extent of awareness of the causes of CD among the study population are displayed in table 9. A large proportion of both the cases and controls reported being unaware of the causes of CD (72.0 % and 94.3 %, respectively). Of the cases, 11.9 % attributed the cause to stress, followed by food intolerance (5.5 %), dairy products (4.6 %), and smoking (4.6 %). In comparison, the highest reported cause among the control population was food intolerance (1.9 %). A small percentage of both groups attributed CD to a bacterial (cases, 4.1 %; controls 0.5 %), viral (cases, 0.9 %; controls 0.5 %) or water (cases, 0.9 %; controls 0.3 %) related origin (table 9).

DISCUSSION

The role of MAP in the aetiology of CD and the potential mode of transmission if an aetiological link exists remains unknown. There is a large body of poorly conducted case control studies with small sample sizes on different aspects of the aetiology of CD including diet, oral contraceptive use, smoking and microorganisms (Ekbom et al 2004;Korelitz et al 1996). To our knowledge, this study is one of the largest case control studies of recently diagnosed CD patients to examine the potential link between proxy measures of exposure to MAP, such as intake of water, pasteurised milk and dairy products, and the risk of CD. This study is also the first to examine the risk factors associated with the various types of CD, as defined by the Vienna classification (Gasche C et al 2000).

No association was found between the intake of water or proxy measures of consumption of MAP contaminated water and dairy products and the risk of CD. Pasteurised milk intake was found to be associated with a decreased risk in some sub types of CD.

Strengths and weaknesses of the study

The study was designed to minimise the common sources of bias that arise from a case control study. Controls were randomly selected from the catchment population of hospitals from which the cases arose, enhancing comparability and reducing selection bias. Furthermore, the exposure information was collected with a standard questionnaire for both cases and controls without revealing the hypothesis under investigation.

The study population for both cases and controls was mainly derived from the Norwich area, which may lead to a potential bias when interpreting the results. The five centres differed

significantly in the univariable analysis. However, the effect of centre was controlled for in the adjusted and multivariable models. Residual bias arising from this could still affect the results of the study despite adjustment of its effect.

The low response rate noted amongst controls may introduce some bias into the sample where cases may be different from controls since the latter were a self-selected group. Furthermore, it is acknowledged that although all controls were matched by gender, the total number of females recruited was higher among controls than cases. This differential response indicates a higher number of male non-responders in the control group which may in turn also introduce a degree of bias into the study population. These sources of bias are unlikely to invalidate the conclusions of the study.

As with all case control studies, it is difficult to determine if the exposures precede or follow infection since exposure was measured after the disease occurred. This is especially pertinent in view of the chronic nature of MAP infection. Therefore, the ability to infer causality is limited by this uncertainty in the temporal nature of any significant association observed.

Due to incomplete data provision, especially for records five years or longer prior to diagnosis, the proportion of postcode data available decreased. This bias limits the ability of this study to investigate the GIS derived water related exposure factors. However, there was no difference in the proportion of missing postcodes among cases and controls.

Recall bias, a shortcoming known to complicate dietary studies, was minimised by collecting information within five years of onset of symptoms. Sub analysis for cases and controls with a date of onset of symptoms within the last two years of the study showed comparable results indicating that recall bias in this study was minimal. Individuals were also asked about their

knowledge of the causative factors of CD. Only a small proportion of both cases and controls were aware of the recognised risk factors or the hypothesis under investigation (MAP through drinking water or milk and dairy products), making it unlikely that this has introduced bias. For the GIS derived data, previously recorded information was used, therefore eliminating recall bias. Confounding factors were also controlled.

The primary risk factors under investigation were drinking water, pasteurised milk and dairy product intake. All other associations observed in this study were not part of the original study hypothesis and therefore any associations seen other than those proposed in the primary hypothesis can arise by chance alone. In such circumstances, further studies are necessary before causality can be inferred.

To ensure the validity of logistic regression analysis at least 10 events (and similarly non-events) per variable should be included in a model (Peduzzi et al. 1996). This study had the required sample size to explore several risk factors. We used a p value of 0.005 to indicate statistically significant results in the analysis of the GIS derived data to account for multiple testing.

A potential limitation of the GIS derived variables and the reports of Johne's disease is that they were all measured at an ecological level introducing the potential for ecological bias. We used the smallest geographical unit available i.e. WSZ to minimise this bias, although some WSZs may have had identical GIS variables. Approximately 1 % of participants were excluded from the analyses due to missing data from water companies on WSZs. This proportion is small and thus has a negligible impact on the results of the analyses. In addition, the occurrence of Johne's disease is based on voluntary reporting to the Veterinary Laboratories Agency and maybe incomplete.

During data interpretation it was assumed that all cases contracted the disease as a result of exposure to MAP within the last 10 years. The duration of the latent period between infection with MAP and the development of CD is not known (Reilly et al. 1986). Estimates vary from a few months to several years. We were guided in our choice of 10 years by data on the observed incubation period of Johne's disease. MAP is transmitted in the faeces of infected adults to young animals, where the incubation period is inversely related to the size of the challenge dose, but this could be a prolonged period before the development of clinical disease. Johne's disease is normally only observed in cattle between two and six years of age although the range is from four months to 15 years (Caldow G et al. 2000). Therefore, the GIS derived proxy measures of exposure for this study were obtained for four different time periods ranging from the year before the onset of symptoms to 10 years previously to ensure that a delayed latent period is detected. This study may not detect an association if the incubation period is longer than 10 years.

A number of approaches are available to investigate the hypothesis that MAP plays a role in the aetiology of CD, including case control, cohort or intervention studies. A randomised controlled trial is the preferred approach to test the hypothesis that active MAP infection is the cause of CD. However, this type of study would not determine whether MAP is responsible for the initial insult that triggers subsequent development of CD in the absence of ongoing infection. For this, an epidemiological study is required to investigate previous exposure, ideally using a cohort design with appropriate microbiological detection of MAP in the environment and tissues/body fluids of cases and controls. However, the cost of such an approach is prohibitive and the microbiological tools required to either culture or detect MAP using molecular methods are currently not available. Clearly, the most feasible and cost effective approach to investigate the role of MAP is to use a case control study based on proxy measures of exposure, as in the current study.

Water and MAP

There was no significant association between any of the measures of water intake and CD after controlling for the effect of confounding factors, although in the univariable analysis drinking unboiled water was associated with an increased risk. This may be due to the absence of an aetiological agent in drinking water or because the quantity of the organism is not determined by the volume of water consumed, but rather depends on other factors, such as the number of infected animals defecating on source water and the quality of water treatment.

To support the data obtained from questionnaires, the risk of various water treatment measures, including a comparison of different types of treatment, surface and ground water and animal density were examined. None of the water treatment variables were statistically significant. The absence of an effect for surface versus ground water, animal density and Johne's disease reports may be explained if MAP is ubiquitous in water at levels below the detection limits of current assays. Therefore its presence in source water may not be a good predictor of disease.

Milk and dairy products and MAP

Components of the diet which can be potentially contaminated with MAP were examined. The intake of pasteurised milk was shown to be protective against the development of CD. This is the first observation of a reduced risk associated with pasteurised milk. This effect was shown in cases with penetrating CD and disease of the ileocolon, and is in contrast to the results of a recent case control study where no association was found between the intake of pasteurised milk and the risk of CD (Bernstein et al. 2004).

Dietary factors were also analysed and demonstrated a significant association between the consumption of low amounts of fruits and high intake of meat and the risk of CD. Previous studies have reported a reduced risk associated with fruit intake (Reif et al. 1997;Russel et al. 1998a), lending support to the findings in this study. The significant association observed with eating meat may be attributed to the higher protein intake as identified previously (Shoda et al. 1996). The observation of other significant dietary associations with fish and fruit intake supports this possibility. Nonetheless, in view of the conflicting reduction in risk associated with pasteurised milk, protein intake may not explain the observation. Alternatively, meat may be contaminated with MAP (Rossiter et al. 2001), leading to an increased risk of infection and subsequent development of CD. However, this finding may have arisen by chance alone.

Cod liver oil intake was shown to reduce risk in cases with CD of the ileocolon, an observation which has been reported previously (Simopoulos, 2002). Several case control studies (Riordan et al. 1998) have suggested that sugar intake may contribute to CD, although conflicting findings have been reported. Sugar intake as a risk factor was not investigated in the current study since contamination with MAP is unlikely.

An association between diet and CD may be attributed to the presence of detergent and or emulsifiers in food causing increased intestinal permeability (J Rhodes – personal communication). Dish washing practice was recorded and no association with CD was found. The observation of both a significant positive and negative association for different dietary components in this study does not support this hypothesis, as both are likely to contain particles and detergents.

There is an abundance of literature which supports the role of elemental diets (special liquid diets) (Forbes 2002) in maintaining remission among CD patients, indicating a role for dietary antigens in reducing intestinal inflammation and permeability. Enteral feeding (delivery of liquid

feedings through a tube) has been used to improve CD patients clinically, endoscopically and biochemically (Zachos et al. 2001). Dietary antigens are abundant in the gut and vary by geographical area; this suggests diet represents a plausible explanation for the variation in the occurrence of CD globally, with the observed changes in incidence ascribed to migration and changes in diet.

Confounding factors and CD

The identification of a significant increase in risk with previously recognised risk factors, such as family history (Korelitz et al 1996) (supported by the identification of the CARD15 mutation increasing risk, Hampe et al. 2001) and smoking (Calkins 1989;Katschinski et al. 1993;Russel et al. 1998b;Sandler et al. 1992), is an important observation and lends support to the validity of the findings in this study.

Information was collected regarding other factors that are known or hypothesised to increase the risk of CD. A significant association with previous oral contraceptive use and appendectomy was not found. Although these have been observed to increase risk in several studies (Ekbom et al 2004;Godet et al. 1995;Korelitz et al 1996;Timmer 2003), many of these are small and did not control for other confounding factors. Other authors have also found no association between oral contraceptive use and CD (Lashner et al. 1989).

Surprisingly a history of holidays outside the UK was shown to be associated with a significant reduction in the risk of CD after controlling for confounding. Individuals with CD were less likely to have travelled in the year before they developed symptoms, possibly because they may already have been unwell. Alternatively, a higher risk of CD associated with not travelling abroad may exist, since MAP may be more common in the UK compared with other European

countries (most individuals reported visiting Europe and many controls had lived in Europe compared with none of the cases).

Sub group analysis

It has been hypothesised that CD consists of various diseases related to different genotypes and aetiological factors (Gasche et al 2005). MAP has been specifically associated with fistularising (penetrating) disease (Greenstein 2003). The risk factors for each type of CD were investigated. Family history consistently showed an increased risk in all types of CD and at all sites of the disease. However, other risk factors varied by disease type and this may indicate differences in the aetiology. For instance, smoking was associated with an increased risk of CD affecting the ileocolon, consistent with previous reports (Russel et al 1998b), but is not a risk factor for penetrating disease. Oral contraceptive use was found to be significantly associated with an increased risk of upper GI disease, although it did not affect the risk in the other categories. Intake of meat was only associated with non-stricturing non-penetrating CD and ileocolon disease.

Although our study size is large, the sub group analysis is limited by the small number of cases per group and in some instances, such as with patients with terminal ileum disease, very few cases were available for analysis. Nonetheless, despite this limitation we were still able to identify the important risk factors in the multivariable model.

Interpretation of the results and implications for further research

Despite the absence of an association in the multivariable models between water intake and water treatment and the risk of CD, this study does not exclude the possibility that microorganisms may play a role in the aetiology of CD. There are several factors supporting the suggestion that microbes may contribute to CD and comprehensive reviews of the evidence for the role of MAP in CD have been published (Scientific Committee on Animal Health and Animal Welfare 2000). Nevertheless, the findings that many of the proposed vehicles for MAP transmission to humans (drinking water, pasteurised milk and contact with farm animals) were either not associated or negatively associated with disease, would suggest that MAP may not be an important aetiological agent for CD. The association between meat consumption and some forms of CD is compatible with the MAP hypothesis, though other explanations that do not include MAP exist. As noted previously, further studies are required to refute or confirm this observation.

The results of a large randomised controlled trial using clarithromycin and rifabutin to treat CD in Australia (Selby et al. 2001) have been reported. This study, which aimed to determine if antimycobacterial therapy resulted in long term benefits for CD patient's, presented evidence which refutes a role for MAP infection in the aetiology of CD (Selby WS et al. 2005). It is unlikely that additional case control studies will enhance our knowledge, but rather cohort type studies with collaboration between veterinary clinicians, gastroenterologists, microbiologists and epidemiologists may be beneficial as soon as more reliable methods for the detection of MAP in both the environment and in food are available.

CONCLUSION

The findings of this study do not support a role for either of the two primary hypotheses (drinking water or dairy products) for the transmission of MAP from animals to humans in the aetiology of CD. Indeed this study represents the first observation of a negative association between higher levels of pasteurised milk intake and the risk of CD, an effect seen in patients with penetrating CD and disease of the ileocolon. In addition to the primary hypotheses, an association with farm visits or contact with farm animals was not observed. Taken together these findings do not support a causative role for MAP in the aetiology of CD. An observation which is consistent with the recently reported conclusions of a randomised double blind control trial of antimycobacterial therapy for the treatment of CD.

As expected from earlier studies, family history and smoking were both associated with increased risk in the final model. A negative association with fruit consumption and travel abroad and a positive association with meat intake were also observed in the final model. None of these variables were included as a primary hypothesis and so caution in their interpretation is required. However, the negative association with fruit consumption has been reported previously in more than one study, lending weight to this finding. The negative association with travel abroad was not anticipated and at present remains unexplained. To our knowledge this is the first study to identify an association with meat consumption and, as this was an unexpected finding, no definite conclusions can be drawn. Further work is therefore warranted to confirm causality since unexpected findings such as this can arise by chance. If the association with meat consumption is confirmed, this may in fact be related to total animal protein intake, a finding that has already been reported previously.

References

Andersen, J., Lisby, J. G., Engbaek, K., Ostergaard, T. O., Kornum, L. S., & Binder, V. 1997, "[Mycobacterium paratuberculosis--an etiological agent in Crohn disease?]", *Ugeskr.Laeger*, vol. 159, no. 2, pp. 159-163.

Belluzzi, A., Brignola, C., Campieri, M., Pera, A., Boschi, S., & Miglioli, M. 1996, "Effect of an enteric-coated fish-oil preparation on relapses in Crohn's disease", *N.Engl.J Med*, vol. 334, no. 24, pp. 1557-1560.

Bernstein, C. N., Blanchard, J. F., Rawsthorne, P., & Collins, M. T. 2004, "Population-based case control study of seroprevalence of Mycobacterium paratuberculosis in patients with Crohn's disease and ulcerative colitis", *J.Clin Microbiol.*, vol. 42, no. 3, pp. 1129-1135.

Bingham, S. A., Gill, C., Welch, A., Cassidy, A., Runswick, S. A., Oakes, S., Lubin, R., Thurnham, D. I., Key, T. J., Roe, L., Khaw, K. T., & Day, N. E. 1997, "Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers", *Int.J Epidemiol.*, vol. 26 Suppl 1, p. S137-S151.

Bull, T. J., McMinn, E. J., Sidi-Boumedine, K., Skull, A., Durkin, D., Neild, P., Rhodes, G., Pickup, R., & Hermon-Taylor, J. 2003, "Detection and verification of Mycobacterium avium subsp. paratuberculosis in fresh ileocolonic mucosal biopsy specimens from individuals with and without Crohn's disease", *J.Clin Microbiol.*, vol. 41, no. 7, pp. 2915-2923.

Caldow G & Gunn GJ 2000, *Assessment of surveillance and control of Johne's disease in farm animals in GB*, SAC Veterinary Science Division.

Calkins, B. M. 1989, "A meta-analysis of the role of smoking in inflammatory bowel disease", *Dig.Dis Sci.*, vol. 34, no. 12, pp. 1841-1854.

Chiodini, R. J. & Rossiter, C. A. 1996, "Paratuberculosis: a potential zoonosis?", *Vet.Clin North Am.Food Anim Pract.*, vol. 12, no. 2, pp. 457-467.

Corti, S. & Stephan, R. 2002, "Detection of Mycobacterium avium subspecies paratuberculosis specific IS900 insertion sequences in bulk-tank milk samples obtained from different regions throughout Switzerland", *BMC.Microbiol*, vol. 2, no. 1, p. 15.

Crohn BB, Ginzburg K, & Oppenheimer GD. Regional ileitis: a pathological and clinical entity. J Am Med Assoc 99, 1323-1328. 1932.
Ref Type: Journal (Full)

Dalziel TK. Chronic interstitial enteritis. BMJ , 1068-1070. 1913.
Ref Type: Journal (Full)

Day, N., McKeown, N., Wong, M., Welch, A., & Bingham, S. 2001, "Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium", *Int.J Epidemiol.*, vol. 30, no. 2, pp. 309-317.

Dell'Isola, B., Poyart, C., Goulet, O., Mougenot, J. F., Sadoun-Journo, E., Brousse, N., Schmitz, J., Ricour, C., & Berche, P. 1994, "Detection of Mycobacterium paratuberculosis by polymerase chain reaction in children with Crohn's disease", *J.Infect.Dis*, vol. 169, no. 2, pp. 449-451.

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R., & Kwan, I. 2002, "Increasing response rates to postal questionnaires: systematic review", *BMJ*, vol. 324, no. 7347, p. 1183.

Ekbom, A. & Montgomery, S. M. 2004, "Environmental risk factors (excluding tobacco and microorganisms): critical analysis of old and new hypotheses", *Best.Pract.Res.Clin.Gastroenterol*, vol. 18, no. 3, pp. 497-508.

Ellingson J "Mycobacterium paratuberculosis in US milk", International Association of Food Protection.

Elsaghier, A., Prantera, C., Moreno, C., & Ivanyi, J. 1992, "Antibodies to Mycobacterium paratuberculosis-specific protein antigens in Crohn's disease", *Clin Exp.Immunol.*, vol. 90, no. 3, pp. 503-508.

Forbes, A. 2002, "Review article: Crohn's disease--the role of nutritional therapy", *Aliment.Pharmacol.Ther*, vol. 16 Suppl 4, pp. 48-52.

Gasche C, Scholmerich J, Brynskov J, D'Haens, G., Hanauer, S. B., Irvine E.J., Jewell DP, Rachmilewitz D, Sachar DB, Sandborn WJ, & Sutherland, L. R. 2000, "A simple classification of Crohn's disease: report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998.", *Inflamm.Bowel.Dis*, vol. 6, no. 1, pp. 8-15.

Gasche, C. & Grundtner, P. 2005, "Genotypes and phenotypes in Crohn's disease: do they help in clinical management?", *Gut*, vol. 54, no. 1, pp. 162-167.

Godet, P. G., May, G. R., & Sutherland, L. R. 1995, "Meta-analysis of the role of oral contraceptive agents in inflammatory bowel disease", *Gut*, vol. 37, no. 5, pp. 668-673.

Grant IR 1997, "Zoonotic potential of Mycobaterium paratuberculosis," in *Modern perspectives on Zoonoses*, Holland CV, ed., Dublin Royal Irish Academy, pp. 75-83.

Grant, I. R. 2005, "Zoonotic potential of Mycobacterium avium ssp. paratuberculosis: the current position", *J Appl.Microbiol.*, vol. 98, no. 6, pp. 1282-1293.

Grant, I. R. 2003, "Mycobacterium paratuberculosis and milk", *Acta Vet.Scand.*, vol. 44, no. 3-4, pp. 261-266.

Grant, I. R. 1998, "Does Mycobacterium paratuberculosis survive current pasteurization conditions?", *Appl.Environ.Microbiol*, vol. 64, no. 7, pp. 2760-2761.

Grant, I. R., Ball, H. J., Neill, S. D., & Rowe, M. T. 1996, "Inactivation of Mycobacterium paratuberculosis in cows' milk at pasteurization temperatures", *Appl.Environ.Microbiol.*, vol. 62, no. 2, pp. 631-636.

Grant, I. R., Hitchings, E. I., McCartney, A., Ferguson, F., & Rowe, M. T. 2002, "Effect of commercial-scale high-temperature, short-time pasteurization on the viability of

Mycobacterium paratuberculosis in naturally infected cows' milk", *Appl.Environ.Microbiol.*, vol. 68, no. 2, pp. 602-607.

Grant, I. R., O'Riordan, L. M., Ball, H. J., & Rowe, M. T. 2001, "Incidence of Mycobacterium paratuberculosis in raw sheep and goats' milk in England, Wales and Northern Ireland", *Vet.Microbiol.*, vol. 79, no. 2, pp. 123-131.

Greenstein, R. J. 2003, "Is Crohn's disease caused by a mycobacterium? Comparisons with leprosy, tuberculosis, and Johne's disease", *Lancet Infect.Dis*, vol. 3, no. 8, pp. 507-514.

Hampe, J., Cuthbert, A., Croucher, P. J., Mirza, M. M., Mascheretti, S., Fisher, S., Frenzel, H., King, K., Hasselmeyer, A., MacPherson, A. J., Bridger, S., van, D. S., Forbes, A., Nikolaus, S., Lennard-Jones, J. E., Foelsch, U. R., Krawczak, M., Lewis, C., Schreiber, S., & Mathew, C. G. 2001, "Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations", *Lancet*, vol. 357, no. 9272, pp. 1925-1928.

Hawthorne, A. B., Daneshmend, T. K., Hawkey, C. J., Belluzzi, A., Everitt, S. J., Holmes, G. K., Malkinson, C., Shaheen, M. Z., & Willars, J. E. 1992, "Treatment of ulcerative colitis with fish oil supplementation: a prospective 12 month randomised controlled trial", *Gut*, vol. 33, no. 7, pp. 922-928.

Hermon-Taylor, J., Bull, T. J., Sheridan, J. M., Cheng, J., Stellakis, M. L., & Sumar, N. 2000, "Causation of Crohn's disease by Mycobacterium avium subspecies paratuberculosis", *Can.J.Gastroenterol.*, vol. 14, no. 6, pp. 521-539.

Hudson, M., Chitolie, A., Hutton, R. A., Smith, M. S., Pounder, R. E., & Wakefield, A. J. 1996, "Thrombotic vascular risk factors in inflammatory bowel disease", *Gut*, vol. 38, no. 5, pp. 733-737.

Hugot, J. P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J. P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C. A., Gassull, M., Binder, V., Finkel, Y., Cortot, A., Modigliani, R., Laurent-Puig, P., Gower-Rousseau, C., Macry, J., Colombel, J. F., Sahbatou, M., & Thomas, G. 2001, "Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease", *Nature*, vol. 411, no. 6837, pp. 599-603.

Johne HA & .Frothingham J. Ein eigenthuemlicher fallvon tuberculose beim rind. Dtsch.Z.Tiermed.Pathol. 21, 438-454. 1895.
Ref Type: Journal (Full)

Katschinski, B., Fingerle, D., Scherbaum, B., & Goebell, H. 1993, "Oral contraceptive use and cigarette smoking in Crohn's disease", *Dig.Dis Sci.*, vol. 38, no. 9, pp. 1596-1600.

Korelitz, B. I. & Prantera, C. 1996, *Crohn's disease*, 1 edn, M. Dekker, New York.

Lashner, B. A., Kane, S. V., & Hanauer, S. B. 1989, "Lack of association between oral contraceptive use and Crohn's disease: a community-based matched case-control study", *Gastroenterology*, vol. 97, no. 6, pp. 1442-1447.

Le Dantec, C., Duguet, J. P., Montiel, A., Dumoutier, N., Dubrou, S., & Vincent, V. 2002a, "Chlorine disinfection of atypical mycobacteria isolated from a water distribution system", *Appl.Environ.Microbiol*, vol. 68, no. 3, pp. 1025-1032.

Le Dantec, C., Duguet, J. P., Montiel, A., Dumoutier, N., Dubrou, S., & Vincent, V. 2002b, "Chlorine disinfection of atypical mycobacteria isolated from a water distribution system", *Appl.Environ.Microbiol*, vol. 68, no. 3, pp. 1025-1032.

Le Dantec, C., Duguet, J. P., Montiel, A., Dumoutier, N., Dubrou, S., & Vincent, V. 2002c, "Occurrence of mycobacteria in water treatment lines and in water distribution systems", *Appl.Environ.Microbiol*, vol. 68, no. 11, pp. 5318-5325.

Le Dantec, C., Duguet, J. P., Montiel, A., Dumoutier, N., Dubrou, S., & Vincent, V. 2002d, "Occurrence of mycobacteria in water treatment lines and in water distribution systems", *Appl.Environ.Microbiol*, vol. 68, no. 11, pp. 5318-5325.

Lee J 2002, "The Mycobacterium avium complex and M. avium subsp. paratuberculosis; risks to health and incidence in water".

Lennard-Jones, J. E. 1989, "Classification of inflammatory bowel disease", *Scand.J Gastroenterol Suppl*, vol. 170, pp. 2-6.

Loftus, E. V., Jr. 2004, "Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences", *Gastroenterology*, vol. 126, no. 6, pp. 1504-1517.

Logan, R. F. & Kay, C. R. 1989, "Oral contraception, smoking and inflammatory bowel disease--findings in the Royal College of General Practitioners Oral Contraception Study", *Int.J Epidemiol.*, vol. 18, no. 1, pp. 105-107.

Lund, B. M., Gould, G. W., & Rampling, A. M. 2002, "Pasteurization of milk and the heat resistance of Mycobacterium avium subsp. paratuberculosis: a critical review of the data", *Int.J.Food Microbiol.*, vol. 77, no. 1-2, pp. 135-145.

Meylan, M., Rings, D. M., Shulaw, W. P., Kowalski, J. J., Bech-Nielsen, S., & Hoffsis, G. F. 1996, "Survival of Mycobacterium paratuberculosis and preservation of immunoglobulin G in bovine colostrum under experimental conditions simulating pasteurization", *Am J Vet.Res.*, vol. 57, no. 11, pp. 1580-1585.

Mishina, D., Katsel, P., Brown, S. T., Gilberts, E. C., & Greenstein, R. J. 1996, "On the etiology of Crohn disease", *Proc.Natl.Acad.Sci.U.S.A*, vol. 93, no. 18, pp. 9816-9820.

Moss, M. T., Sanderson, J. D., Tizard, M. L., Hermon-Taylor, J., El Zaatari, F. A., Markesich, D. C., & Graham, D. Y. 1992, "Polymerase chain reaction detection of Mycobacterium paratuberculosis and Mycobacterium avium subsp silvaticum in long term cultures from Crohn's disease and control tissues", *Gut*, vol. 33, no. 9, pp. 1209-1213.

Naser, S. A., Ghobrial, G., Romero, C., & Valentine, J. F. 2004, "Culture of Mycobacterium avium subspecies paratuberculosis from the blood of patients with Crohn's disease", *Lancet*, vol. 364, no. 9439, pp. 1039-1044.

Naser, S. A., Hulten, K., Shafran, I., Graham, D. Y., & El Zaatari, F. A. 2000, "Specific seroreactivity of Crohn's disease patients against p35 and p36 antigens of M. avium subsp. paratuberculosis", *Vet.Microbiol.*, vol. 77, no. 3-4, pp. 497-504.

Pavlik, I., Bartl, J., Dvorska, L., Svastova, P., du, M. R., Machackova, M., Yayo, A. W., & Horvathova, A. 2000, "Epidemiology of paratuberculosis in wild ruminants studied by

restriction fragment length polymorphism in the Czech Republic during the period 1995-1998", *Vet.Microbiol.*, vol. 77, no. 3-4, pp. 231-251.

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. 1996, "A simulation study of the number of events per variable in logistic regression analysis", *J Clin.Epidemiol.*, vol. 49, no. 12, pp. 1373-1379.

Reif, S., Klein, I., Lubin, F., Farbstein, M., Hallak, A., & Gilat, T. 1997, "Pre-illness dietary factors in inflammatory bowel disease", *Gut*, vol. 40, no. 6, pp. 754-760.

Reilly, R. P. & Robinson, T. J. 1986, "Crohn's disease--is there a long latent period?", *Postgrad.Med.J.*, vol. 62, no. 727, pp. 353-354.

Riordan, A. M., Ruxton, C. H., & Hunter, J. O. 1998, "A review of associations between Crohn's disease and consumption of sugars", *Eur.J Clin.Nutr.*, vol. 52, no. 4, pp. 229-238.

Rossiter, C. A. & Henning, W. R. 2001, "Isolation of M paratuberculosis from thin market cows at slaughter", *J Animal Science*, vol. 79 Supplement 1, pp. 113-114.

Russel, M. G., Engels, L. G., Muris, J. W., Limonard, C. B., Volovics, A., Brummer, R. J., & Stockbrugger, R. W. 1998a, "Modern life' in the epidemiology of inflammatory bowel disease: a case-control study with special emphasis on nutritional factors", *Eur.J Gastroenterol Hepatol*, vol. 10, no. 3, pp. 243-249.

Russel, M. G., Volovics, A., Schoon, E. J., van Wijlick, E. H., Logan, R. F., Shivananda, S., & Stockbrugger, R. W. 1998b, "Inflammatory bowel disease: is there any relation between smoking status and disease presentation? European Collaborative IBD Study Group", *Inflamm.Bowel.Dis*, vol. 4, no. 3, pp. 182-186.

Sandler, R. S., Wurzelmann, J. I., & Lyles, C. M. 1992, "Oral contraceptive use and the risk of inflammatory bowel disease", *Epidemiology*, vol. 3, no. 4, pp. 374-378.

Scientific Committee on Animal Health and Animal Welfare 2000, *Possible links between Crohn's disease and Paratuberculosis*, European Commission, SANCO/B3/R16/2000.

Selby WS, Crotty, B., Florin, T., & Pavli, P. 2005, *Antimycobacterial therapy: Past it or present? Results of the 2004 Australian Trial – Balanced Overview*. Birmingham.

Selby, W., Crotty, B., Florin, T., & Pavli, P. 2001, "Combination anti-paratuberculosis therapy with clarithromycin, rifabutin and clofazimine in the treatment of Crohn's disease. A preliminary report on the Australian multicentre trial", *J.Gastroenterol.Hepatol*, vol. 16 Suppl, no. A8.

Shoda, R., Matsueda, K., Yamato, S., & Umeda, N. 1996, "Epidemiologic analysis of Crohn disease in Japan: increased dietary intake of n-6 polyunsaturated fatty acids and animal protein relates to the increased incidence of Crohn disease in Japan", *Am.J.Clin.Nutr.*, vol. 63, no. 5, pp. 741-745.

Stenson, W. F., Cort, D., Rodgers, J., Burakoff, R., DeSchryver-Kecskemeti, K., Gramlich, T. L., & Beeken, W. 1992, "Dietary supplementation with fish oil in ulcerative colitis", *Ann.Intern.Med*, vol. 116, no. 8, pp. 609-614.

Streeter, R. N., Hoffsis, G. F., Bech-Nielsen, S., Shulaw, W. P., & Rings, D. M. 1995, "Isolation of Mycobacterium paratuberculosis from colostrum and milk of subclinically infected cows", *Am.J.Vet.Res.*, vol. 56, no. 10, pp. 1322-1324.

Sung, N. & Collins, M. T. 1998, "Thermal tolerance of Mycobacterium paratuberculosis", *Appl.Environ.Microbiol*, vol. 64, no. 3, pp. 999-1005.

Taylor, T. K., Wilks, C. R., & McQueen, D. S. 1981, "Isolation of Mycobacterium paratuberculosis from the milk of a cow with Johne's disease", *Vet.Rec.*, vol. 109, no. 24, pp. 532-533.

Thompson, D. E. 1994, "The role of mycobacteria in Crohn's disease", *J.Med.Microbiol.*, vol. 41, no. 2, pp. 74-94.

Timmer, A. 2003, "Environmental influences on inflammatory bowel disease manifestations. Lessons from epidemiology", *Dig.Dis*, vol. 21, no. 2, pp. 91-104.

Travis, S. P. 1995, "Mycobacteria on trial: guilty or innocent in the pathogenesis of Crohn's disease?", *Eur.J.Gastroenterol.Hepatol.*, vol. 7, no. 12, pp. 1173-1176.

Van Kruiningen, H. J. 1999, "Lack of support for a common etiology in Johne's disease of animals and Crohn's disease in humans", *Inflamm.Bowel.Dis*, vol. 5, no. 3, pp. 183-191.

Vessey, M., Jewell, D., Smith, A., Yeates, D., & McPherson, K. 1986, "Chronic inflammatory bowel disease, cigarette smoking, and use of oral contraceptives: findings in a large cohort study of women of childbearing age", *Br.Med J (Clin.Res.Ed)*, vol. 292, no. 6528, pp. 1101-1103.

Wendt, S. L., George, K. L., Parker, B. C., Gruft, H., & Falkinham, J. O., III 1980, "Epidemiology of infection by nontuberculous Mycobacteria. III. Isolation of potentially pathogenic mycobacteria from aerosols", *Am Rev.Respir Dis*, vol. 122, no. 2, pp. 259-263.

Whan, L. B., Grant, I. R., Ball, H. J., Scott, R., & Rowe, M. T. 2001, "Bactericidal effect of chlorine on Mycobacterium paratuberculosis in drinking water", *Lett.Appl.Microbiol.*, vol. 33, no. 3, pp. 227-231.

Willett, W. C., Reynolds, R. D., Cottrell-Hoehner, S., Sampson, L., & Browne, M. L. 1987, "Validation of a semi-quantitative food frequency questionnaire: comparison with a 1-year diet record", *J Am Diet.Assoc*, vol. 87, no. 1, pp. 43-47.

Willett, W. 1998, *Nutritional epidemiology*, 2nd ed edn, Oxford University Press, New York.

Zachos, M., Tondeur, M., & Griffiths, A. M. 2001, "Enteral nutritional therapy for inducing remission of Crohn's disease", *Cochrane.Database.Syst.Rev.* no. 3, p. CD000542.

APPENDIX 1 EXAMPLE OF THE DIAGNOSTIC CRITERIA

Patient number ☐☐☐☐☐☐

Consultant name ……

Year of diagnosis ……

DIAGNOSTIC CRITERIA FOR PATIENTS WITH CROHN'S DISEASE

A. EVIDENCE OF MACROSCOPIC SMALL BOWEL DISEASE

Has patient had small bowel follow-through/small bowel enema?

Yes ☐ No ☐

If **YES**, please complete all the boxes in the table below. If **NO**, please go to **section B**.

REGION	**Affected** (Y/N/NV)	**Mucosal ulceration** (Y/N/NV)	**Stricture** (Y/N/NV)	**Fistula** (Y/N/NV)	**Other** (Please state)
Duodenum					
Jejunum					
Non-terminal Ileum					
Terminal Ileum					

Y = Yes; N = No; NV = Not Visualised

B. EVIDENCE OF MACROSCOPIC LARGE BOWEL DISEASE AT COLONOSCOPY

Has patient had a colonoscopy?

Yes ☐ No ☐

If **YES**, please complete all the boxes in the table below. If **NO**, please go to **section C**.

REGION	**Macroscopic Changes Present** (Y/N/NE)	**Aphthous ulceration** (Y/N/NE)	**Non-aphthous (larger) ulceration** (Y/N/NE)	**Cobble-stoning** (Y/N/NE)	**Other** (Please state)
Rectum					
Sigmoid					
Descending					
Transverse					
Ascending					
Caecum					
Terminal Ileum					

Y = Yes; N = No; NE = Not Examined

At colonoscopy are the ulceration and inflammatory changes:

Discontinuous (skip lesions) ☐ Continuous ☐

C. EVIDENCE OF MACROSCOPIC LARGE BOWEL DISEASE AT BARIUM ENEMA

Has patient had a barium enema?

Yes ☐ No ☐

If **YES**, please complete all boxes in the table below. If **NO**, please go to **section D**.

REGION	**Radiological changes seen** (Y/N/NE)	**Ulceration** (Y/N/NE)	**Stricture** (Y/N/NE)	**Fistula** (Include site) (Y/N/NE)	**Other** (Please state) (Y/N/NE)
Rectum					
Sigmoid					
Descending					
Transverse					
Ascending					
Caecum					

Y = Yes; N = No; NE = Not Examined

At barium enema are the radiological changes:

Discontinuous (skip lesions) ☐ Continuous ☐

D. EVIDENCE OF HISTOLOGICAL CHANGES OF CROHN'S DISEASE

Has the patient had a biopsy showing histological changes of Crohn's disease?

Yes ☐ No ☐

If **YES**, please complete each of the boxes in the table below. If **NO**, please go to **section E**.

REGION	**Has histology been taken?** i.e. at colonoscopy or surgery (Y/N)	**Are non-caeseating granuloma present?** (Y/N/NH /NC)	**Does inflammat-ion spread beyond mucosa?** i.e.biopsies from colonoscopy (Y/N/NH/ NA)	**Is inflammat-ion discon-tinuous?** (Y/N/NH/ NC)	**Are lymphoid aggregates present?** (Y/N/NH/ NC)	**Is colonic mucin present?** (Y/N/NH/ NC)
Small bowel						NA
Large bowel						
Other (Please state)						

Y = Yes; N = No; NH = No Histology; NC = No Comment made; NA = Not able to Access

E. HAS PATIENT HAD ANY OF THE FOLLOWING SURGICAL PROCEDURES?

PROCEDURE	YES / NO
Right hemicolectomy	
Small bowel resection	
Stricturoplasty (Please include number)	
Colectomy (Type)	
Other (Please state)	

F. HAS THE PATIENT EVER HAD AN INTRA-ABDOMINAL ABSCESS?

Yes ☐ No ☐

If **YES**, which site(s) _____

Method of diagnosis (ultra-sound, CT, surgical) _____

Any other region of gastrointestinal tract affected? Yes ☐ No ☐

Region(s) affected e.g. mouth, stomach, perianal disease:

_____

_____

G. EVIDENCE OF MACROSCOPIC LARGE BOWEL DISEASE AT RIGID SIGMOIDOSCOPY

Has patient had a rigid sigmoidoscopy?

Yes ☐ No ☐

If **YES**, please complete all the boxes in the table below. If **NO**, please go to section H.

REGION	**Macroscopic Changes Present** (Y/N/NE)	**Aphthous ulceration** (Y/N/NE)	**Non-aphthous (larger) ulceration** (Y/N/NE)	**Cobble-stoning** (Y/N/NE)	**Other** (Please state)
Rectum					

H. EVIDENCE OF MACROSCOPIC LARGE BOWEL DISEASE AT FLEXIBLE SIGMOIDOSCOPY

Has patient had a flexible sigmoidoscopy?

Yes ☐ No ☐

If **YES**, please complete all the boxes in the table below. If **NO**, please go to section H.

REGION	**Macroscopic Changes Present** (Y/N/NE)	**Aphthous ulceration** (Y/N/NE)	**Non-aphthous (larger) ulceration** (Y/N/NE)	**Cobble-stoning** (Y/N/NE)	**Other** (Please state)
Rectum					
Sigmoid					
Descending					
Splanchnic Flexur					

APPENDIX 2 EXAMPLE OF THE STUDY QUESTIONNAIRE

Study number ☐☐☐☐☐☐

RISK FACTORS FOR CROHN'S DISEASE

Questionnaire

Please read the information sheet enclosed with this questionnaire

This questionnaire asks about present and previous addresses and about what you eat and drink. Please answer as many of the questions as you can. There are no right or wrong answers. It is OK to answer "don't know". If you need any help please feel free to contact:

.....

Your answers will be kept strictly confidential

Thank you for your help with this research

YOUR PERSONAL DETAILS

1. Sex (Please tick) ☐ Male ☐ Female

2. Date of Birth Day ☐☐ Month ☐☐ Year 1 9 ☐☐ Age ☐☐ years

3. Have you ever been diagnosed with Crohn's disease? (Please tick)
   ☐ Yes ☐ No

4. What is your current home address?

   _____

   _____

   _____

   Postcode

   When did you move to this address? Month ☐☐ Year ☐☐☐☐

5. Please list any other addresses that you have lived at for at least 3 months over the last 10 years, starting with the most recent. (Please include term-time university and college addresses).

   Previous address 1

   _____

   _____

   _____

   Postcode

   Date moved to this address Month ☐☐ Year ☐☐☐☐

   Date left this address Month ☐☐ Year ☐☐☐☐

Previous address 2

Postcode

Date moved to this address Month Year

Date left this address Month Year

Previous address 3

Postcode

Date moved to this address Month Year

Date left this address Month Year

Previous address 4

Postcode

Date moved to this address Month Year

Date left this address Month Year

Previous address 5

Postcode

Date moved to this address Month Year

Date left this address Month Year

Previous address 6

_____

_____

_____

Postcode

Date moved to this address Month ☐☐ Year ☐☐☐☐

Date left this address Month ☐☐ Year ☐☐☐☐

Questions 6-8 are about your water supply in the year _____. We will compare your answers with a patient who developed Crohn's disease at this time.

6. Which best describes the address where you lived *for the longest* in the year _____? (Please tick)

☐ House/flat/apartment ☐ Residential home

☐ Nursing home ☐ Boarding school

☐ Hostel ☐ University/college hall of residence

☐ Other (please state) .....

7. Which one of the following two options best describes the water supply at the address where you lived *for the longest* in the year _____? (Please tick)

☐ Provided by a public water company ☐ Private water supply

If you had a **private supply**, which of the options below best describes it? (Please tick)

☐ Borehole ☐ Stream/river ☐ Pond/lake

☐ Sunken well ☐ Dyke ☐ Reservoir

☐ Ditch ☐ Rainwater tank ☐ Don't know

☐ Spring ☐ Other (please specify) .....

8. Did you use a water filter at the address you lived at *for the longest* in the year _____? (Please tick)

☐ Yes ☐ No ☐ Don't know

Questions 9-15 are about your main place of work or education and holidays in the specified time periods. We will compare your answers with someone who developed Crohn's disease during this time.

9. Did you work or study at a location outside your home in the year _____? (Please tick)

☐ Yes ☐ No

If **NO**, please go to **question 12**.

If **YES**, what was your occupation? .....

10. How many days per week did you work/study away from home? .....

11. Please provide the name and address of the workplace/college you were at *for the longest* in the year _____.

Name _____

Address _____

_____

_____

Postcode _____

When did you start working/attending there? Month ☐☐ Year ☐☐☐☐

12. Please provide the name and address of your main place of work/study you were at *for the longest* in the year **1993**.

Name _____

Address _____

_____

_____

Postcode _____

13. Between the years _____ - _____, did you travel to any countries outside the UK for 7 days or longer? (Please tick)

☐ Yes ☐ No ☐ Don't know

If **YES**, please list the countries below.

Country	Month and year of visit	Length of stay (days)

14. In the year _____, did you have any holidays on a farm? (Please tick)

☐ Yes ☐ No ☐ Don't know

If **YES**, how many holidays in that year? ☐☐ holidays

What was the average length of each holiday? ☐☐ days

15. In the year _____, did you visit farms for any reason (excluding any farm holidays)? (Please tick)

☐ Yes ☐ No ☐ Don't know

If **YES**, approximately how many times in that year? ☐☐ times

Did you have any contact with farm animals?

☐ Yes ☐ No ☐ Don't know

If **YES**, what farm animals? .....

DRINKS SECTION

The next section is about how often you consumed the following drinks in the year _____. We will compare your answers with someone who developed Crohn's disease during this year. For most drinks there is an amount shown in a common household unit e.g. cup, teaspoonful. Please put a tick (✔) on every line to indicate how often, **on average**, you drank the specified amount in the year _____.

Example

For pasteurised milk, the amount is a glass. If you drank 2 glasses of milk a day in the year _____, you should put a tick in the column headed "2-3 per day", as shown below:

Drink	None/ <1 per month	1-3 per month	1 per week	2-4 per week	5-6 per week	1 per day	2-3 per day	4-5 per day	6+ per day
Pasteurised milk (glass)							✔		

16. Please put a tick (✔) on every line below. All the questions are about how often, **on average**, you drank the following drinks in the year _____.

Drink	None/ <1 per month	1-3 per month	1 per week	2-4 per week	5-6 per week	1 per day	2-3 per day	4-5 per day	6+ per day
Pasteurised milk (glass)									
Pasteurised milk added to tea or coffee									
Pasteurised milk on breakfast cereals									
Pasteurised milky drinks e.g. hot chocolate, horlicks, milk shake (cup)									
Dried milk powder (1 teaspoon)									
Boiled drinking water from your HOME supply e.g. in tea, coffee (cup)									
Unboiled drinking water from your HOME supply e.g. in cold drinks, squash (glass)									
Bottled water (glass)									
Boiled drinking water from your WORK or COLLEGE supply e.g. in tea, coffee (cup)									
Unboiled water from your WORK or COLLEGE supply e.g. in cold drinks, squash (glass)									
Fruit juice e.g. orange, apple juice (glass)									
Fruit squash or cordial (glass)									

Please check you have put a tick on every line

17. What type of pasteurised milk did you drink most often in the year _____? (Please tick)

☐ None ☐ Full fat ☐ Semi-skimmed

☐ Skimmed ☐ Dried ☐ Other, please specify .....

18. If you ticked the box for bottled water (in question 16), then please list any brands you drank.

.....

FOOD SECTION

The next section is about how often you ate the following foods in the year _____. We will compare your answers with someone who developed Crohn's disease during this year. For all foods there is an amount described e.g. teaspoonful, carton. Please put a tick (✔) on every line to indicate how often, **on average**, you ate the specified amount in the year _____.

Example

For low fat yoghurt, the amount is a carton. If you ate 3 cartons each week, please put a tick in the column headed "2-4 per week", as shown below:

Food	None/ <1 per month	1-3 per month	1 per week	2-4 per week	5-6 per week	1 per day	2-3 per day	4-5 per day	6+ per day
Low fat yoghurt (carton)				✔					

19. Please put a tick (✔) on each line to show how often, **on average**, you ate the following foods in the year _____.

Food	None/ <1 per month	1-3 per month	1 per week	2-4 per week	5-6 per week	1 per day	2-3 per day	4-5 per day	6+ per day
Apples (1 fruit)									
Bananas (1 fruit)									
Peas: fresh, frozen or tinned (medium serving)									
Single or sour cream (tablespoon)									
Double or clotted cream (tablespoon)									
Low fat yoghurt (carton)									
Full fat yoghurt (carton)									
Hard cheeses e.g. cheddar, edam (medium serving)									
Soft cheeses e.g. cottage cheese, brie (medium serving)									
Ice cream (bowl, cone)									
Butter on bread, toast or sandwiches (pat or teaspoon)									
Butter on jacket potatoes or other vegetables (pat or teaspoon)									

Please check you have put a tick on every line

Food Section continued. **Please tick every line.**

Food	None/ <1 per month	1-3 per month	1 per week	2-4 per week	5-6 per week	1 per day	2-3 per day	4-5 per day	6+ per day
Margarine on bread, toast or sandwiches (pat or teaspoon)									
Margarine on jacket potatoes or other vegetables (pat or teaspoon)									
Beef: roast, steak, mince, stew or casserole (medium serving)									
Canned meat e.g. corned beef (3 slices)									
Fried fish in batter e.g. fish & chips (medium serving)									
Fish fingers, fish cakes (medium serving)									
Other white fish e.g. cod, haddock, plaice, sole, halibut (medium serving)									
Oily fish: fresh or canned e.g. mackerel, kippers, salmon, tuna (medium serving)									
Cod liver oil capsules (single dose)									

Please check you have put a tick on every line

The following questions are about other factors that may increase or decrease the risk of developing Crohn's disease.

20. Have you ever smoked at least 1 cigarette a day for a year or more? (Please tick)

☐ Yes ☐ No

21. Did you smoke cigarettes in the year _____? (Please tick)

☐ Yes ☐ No ☐ Don't know

If **YES**, how many cigarettes did you smoke **each day**? .....

22. Have you had your appendix removed? (Please tick)

☐ Yes ☐ No ☐ Don't know

If **YES**, in what year was it removed? ☐☐☐☐

23. How did you usually wash your dishes in the year _____? (Please tick)

☐ By hand and dried
☐ By hand, rinsed and left to dry
☐ By hand and left to dry
☐ Machine washed
☐ Other, please describe .....

24. For women only - did you ever use the oral contraceptive pill before the year _____? (Please tick)

☐ Yes ☐ No

25. Do any of your relatives suffer with Crohn's disease? (Please tick)

☐ Yes ☐ No ☐ Don't know

If **YES**, what is their relationship to you? .....

26. Are you aware of any possible causes of Crohn's disease?

☐ Yes ☐ No

If **YES**, please list these causes .....

This is the end of the questionnaire

Thank you very much for your help with this research

Please return completed questionnaires to:

.....

in the enclosed stamped addressed envelope

APPENDIX 3 THE VIENNA CLASSIFICATION OF CROHN'S DISEASE

(FROM GASCHE AND GRUNDTNER, 2005)

Type

Stricturing

- Constant luminal narrowing (identified by radiological, endoscopic or surgical procedures).
- Prestenotic dilatation or obstructive signs/symptoms.
- No presence of penetrating disease at any time.

Penetrating

- Intra-abdominal or perianal fistulas at any time.
- Inflammatory masses and/or abscesses at any time.
- Perianal ulcers at any time.
- Post-operative intra-abdominal complications and perianal skin tags are excluded.

Non-stricturing non-penetrating

- Inflammatory disease which has never been complicated at any time.

Location

Upper GI

- Any disease location proximal to the terminal ileum, despite additional involvement of the terminal ileum or colon.
- Excludes the mouth.

Terminal ileum

- Limited to the terminal ileum[2] with or without involvement of the caecum.

Ileocolon

- Disease of the terminal ileum with or without involvement of the caecum and any site between the ascending colon and rectum.

Colon

- Any site between the caecum and rectum.
- No involvement of the upper GI or small bowel.

[2] Lower third of the small bowel

APPENDIX 4 FOOD AND DRINK CATEGORIES SHOWING AVERAGE WEIGHTS FOR VOLUMES AND PORTIONS

Category	Food or drink item	Average portion size	Weight/volume
Boiled water	Boiled drinking water at home e.g tea	Cup	200 mls
	Boiled drinking water at work e.g tea	Cup	200 mls
Unboiled water	Unboiled drinking water at home	Glass	200 mls
	Unboiled drinking water at work	Glass	200 mls
Bottled water	Bottled water	Glass	200 mls
Pasteurised milk	Pasteurised milk	Glass	200 mls
	Pasteurised milk added to tea or milk		5 mls
	Pasteurised milk on breakfast cereal	Bowl	100 mls
	Pasteurised milky drinks e.g. horlicks	Cup	190 mls
	Dried milk powder	Teaspoon	5 mls
Fruit	Fruit juice	Glass	200 mls
	Apples	1 fruit	130 grams
	Banana	1 fruit	148 grams
	Peas: fresh, frozen or tinned	Medium serving	105 grams
Dairy Products	Single or sour cream	Tablespoon	15 mls
	Double or clotted cream	Tablespoon	15 mls
	Low fat yoghurt	Carton	125 grams
	Full fat yoghurt	Carton	125 grams
	Hard cheese eg. chedder	Medium serving	60 grams
	Soft cheese eg. brie	Medium serving	46 grams
	Ice cream	Bowl or cone	98 grams
	Butter on bread, toast or sandwich	Pat or teaspoon	16 grams
	Butter on jacket potatoes or vegetables	Pat or teaspoon	16 grams
	Margarine on bread, toast or sandwich	Pat or teaspoon	16 grams
	Margarine on jacket potatoes or vegetables	Pat or teaspoon	16 grams
Meat	Beef: roast, steak, mince, stew or casserole	Medium serving	100 grams
	Canned meat e.g. corned beef	3 slices	198 grams
Fish	Fried fish in batter .e.g. fish and chips	Medium serving	180 grams
	Fish fingers or cakes	Medium serving	155 grams
	Other white fish e.g. cod	Medium serving	155 grams
	Oily Fish: fresh or canned e.g. mackerel	Medium serving	100 grams
Cod liver oil	Cod liver oil	Capsules	Single dose