Food Standards Scotland
foodstandards.gov.scot

1495 UNIVERSITY OF ABERDEEN

# Factors Affecting Variations in Campylobacter Disease Rates in Scotland

February 2020

s

School of Biological Sciences
University of Aberdeen
Cruickshank Building

St Machar Drive
Aberdeen AB23 8GN
United Kingdom

Tel: +44 (0)1224 272699

Email: n.strachan@abdn.ac.uk

Food Standards Scotland

Reference FS101106

# Final Report:

# Findings

# January 2015 – January 2019

# Contents

# Figures

# Tables

# Annexes (in Report Part 2)

# Lay Summary

## The Background

*Campylobacter* is the main cause of bacterial gastroenteritis in the UK. In Scotland during 2018 there were 6096 reported cases of human campylobacteriosis. Previous work has established an apparent lower incidence of reported *Campylobacter* infections in deprived populations but this is not observed in hospitalised cases. It was not clear whether this was actually a true reflection of the disease incidence, an artefact of reporting or a signature of differential health care use by these communities. This study was commissioned by Food Standards Scotland (FSS) to understand why there are differences in disease incidence between more and less deprived populations and to obtain an up to date picture of campylobacteriosis in Scotland.

## The Study

This project investigated the origin of these differences between people from deprived and prosperous areas in four ways:

(1) Investigating potential biases at three different levels of the reporting pyramid: the community level, the GP level and the reported case level.
(2) Analysing retrospective and prospective case and hospitalisation discharge data to determine whether the reported variation in disease has changed.
(3) Carrying out a case-control study to identify the sources of human campylobacteriosis.
(4) Performing a case-case analysis to determine differences in risk factors for deprived and less deprived (affluent) populations.

## The Findings

<u>The study found more campylobacteriosis cases reported in the less deprived areas.</u> In total there remained a 19% excess of campylobacteriosis cases in the less deprived Scottish Index of Multiple Deprivation quintile areas (SIMD2 to SIMD5). This is six percentage points lower than that observed between 2000 and 2006 but is still statistically significant. Investigation of the way these data were reported did not identify a reporting bias therefore the study found that it is likely that this difference is genuine. However, GPs cited recent foreign travel as being a very important consideration when requesting a stool sample and there was some evidence that having prolonged "nausea or vomiting" symptoms was more likely to lead to an individual from a least deprived background making a doctor's appointment. Further, in the case control study, foreign travel was more common in cases from least deprived (SIMD5) compared with most deprived (SIMD1) areas. It was estimated that this might explain around a third of the difference in cases by deprivation.

<u>Those living in the most deprived areas are more likely to be hospitalised with campylobacteriosis.</u> In total there was a 9% excess hospitalisation rate for the

population living in the two most deprived deprivation quintiles (i.e. SIMD1 and SIMD2). This may be attributed to the following factors:

- the high SIMD1 and SIMD2 populations in these areas that are close (<10km) to a hospital and
- lower health status (e.g. coexistent ill health) and/or lower level of socio-economic support.

Some health and behavioural differences were observed between cases from least and most deprived areas. The case-case study found that taking antacids and H2 blockers as well as washing raw chicken was associated with cases from the more deprived areas whilst having a public water supply and cutting up raw chicken was associated with cases from the richest ones.

Hospitalisation rates of campylobacteriosis have trebled for people aged over 65 years since 2005. This is an increasing concern because of Scotland's ageing population. This should be contrasted with hospitalisation rates in children (<15 years) which have remained relatively stable during this period.

The main findings of the case-control study that combined cases and controls from the most (SIMD1) and least (SIMD5) populations were:

- Consumption of some chicken and poultry products were a significant risk factor except for raw chicken handled in the kitchen. Specifically: eating chicken liver pâté prepared at home; eating chicken lightly cooked; eating chicken outside the home (not restaurant, take-away or fast food); eating poultry (other than chicken) at a restaurant were all significant in all of the analysis. However, consumption of chicken is a complex risk factor and can appear to be "protective" depending on how and where it is prepared. For example, counter-intuitively, raw chicken handled in the kitchen was found to be "protective."
- Non-food risk factors were also consistently significant in all of the analysis. Being prescribed PPIs and having white ethnicity increased risk of campylobacteriosis whilst using an indoor swimming pool/toddler pool decreased it.
- Foreign travel was an important risk factor for campylobacteriosis with 24% of cases reporting this. The risk was greatest when travelling to Asia (including Turkey) and lowest when visiting North America.

Carrying out a national case-control study is challenging because of ethical requirements, logistics and low participation rates among both cases and controls. The amount of paperwork and time required to obtain all permissions to carry out the study was very substantial. In the current study response rates were low (22.7% of cases and 10.6% of controls) and future studies will need to address this. However, such studies do never-the-less provide valuable information that has the potential to be acted upon by FSS and community health protection teams.

**The Conclusions**

There remains an excess of campylobacteriosis cases in the least deprived populations of Scotland which is real and not an artefact of the reporting system. A substantial part of this difference is associated with foreign travel which is more common in the least deprived parts of the population. The excess of hospitalisations in the poorer part of the population may in part be explained by areas closer to a hospital tending to be more deprived.

# Abbreviations

95% CIs – 95% Confidence Intervals

AA – Ayrshire and Arran Health Board

ACMSF – Advisory Committee on the Microbiological Safety of Foods

BR - Borders Health Board

CHI - Community Health Index

CPHM – Consultant in Public Health Medicine

DG - Dumfries and Galloway Health Board

eDRIS - Electronic Data Research and Innovation Service at NHS Scotland

FF – Fife Health Board

FSS – Food Standards Scotland

FSAS – Food Standards Agency Scotland

FV – Forth Valley Health Board

GC – Glasgow and Clyde Health Board

GI – Gastrointestinal Infection

GP – General Practice (Medical)

GR – Grampian Health Board

HG – Highland Health Board

HPS – Health Protection Scotland

IID and IID2 – Studies of Intestinal Infectious disease in the UK funded by the Food Standards Agency

ISD - Information Services Division

LN – Lanarkshire Health Board

LO – Lothian Health Board

MCMC - Markov Chain Monte Carlo

MMDLs- Medical microbiological diagnostic laboratories

NHS Scotland – the National Health Service Scotland

NHS/HSC R&D – NHS/HSC Research and Development offices

NRS - National Records of Scotland

OR - Odds Ratio

OR – Orkney Health Board

PAF – Population Attributable Fraction

PBPP - Public Benefit and Privacy Panel

PPI – Proton Pump Inhibitor

PTI – Practice Team Information

PWS – Private Water Supplies

REC – Research Ethics Committee

RCG3 diagnoses – Read Code Grouping (RCG) for 'Gastroenteritis of possible infectious origin'

SH - Shetland Health Board

SIMD – Scottish Index of Multiple Deprivation

SNAP –Survey aNAlysis Package

SPIRE - Scottish Primary Care Information Resource

TY – Tayside Health Board

WI – Western Isles Health Board

# Glossary

**Carstairs score** is a measure of deprivation in Scotland. The score is a measure of access to "those goods and services, resources and amenities and of a physical environment which are customary in society".

**Case-control study** - is an analytical epidemiological method that compares risk factors of people who have been ill (e.g. with campylobacteriosis) with a control group who have not been ill.

**Datazones** (comprise on average 800 people) of which there are approximately 6,500 in Scotland and are the population units upon which SIMD is calculated.

A **Postcode Sector** is the set of unit postcodes that are the same apart from the last two characters (e.g. Postcode AB24 3UU is part of the postcode sector AB24 3).

The **Scottish Index of Multiple Deprivation (SIMD)** defines deprivation "as the range of problems that arise due to lack of resources or opportunities covering health, safety, education, employment, housing and access to services as well as financial aspects".

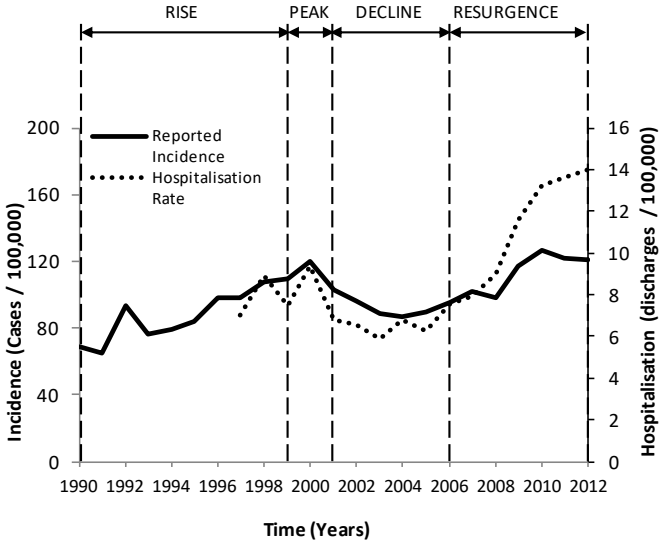
# 1. Introduction

## 1.1 Background

*Campylobacter* is the largest cause of bacterial gastroenteritis in the developed world (Blaser 1997) with 63,000 cases reported in the UK during 2017 (ACMSF 2018) of which 5,796 were reported from Scotland (www.hps.scot.nhs.uk/resourcedocument.aspx?id=6483). Approximately 90% of cases are attributed to *C. jejuni* with most of the remainder to *C. coli* (Gillespie, O'Brien et al. 2002, Roux, Sproston et al. 2013). Since there is significant underreporting the actual number of community cases is likely to be considerably higher (e.g. estimated to be nine-fold higher in the UK (C. C. Tam, Rodrigues et al. 2012)). Further, around 10% of individuals reported as having campylobacteriosis are hospitalised and sequelae include not only severe stomach cramps and diarrhoea but in up to two-thirds of cases musculoskeletal, joint swelling or sensory problems (Zia, Wareing et al. 2003). In the UK it has been reported that *Campylobacter* contributes up to 15% of all Guillain-Barré Syndrome cases (2 for every 10,000 reported campylobacteriosis cases) (C. C. Tam, Rodrigues et al. 2006) and 80 deaths annually (Adak, Long et al. 2002). These cause considerable demands on health services, impose wider economic costs and impacts on those infected and their families and carers.

Figure 1.1 illustrates the 75% increase in reported cases between 1990 and 2012 in Scotland and the 90% increase in hospitalisation rate associated with human campylobacteriosis from the late 1990's to 2012. However, these trends are non-uniform and, in particular, the increase since 2004 is predominantly among the elderly and adult populations (N. J. C. Strachan, Rotariu et al. 2013). This reflects the complex aetiology of this disease. Combining epidemiological methods and microbial source typing has demonstrated that eating chicken is the main UK source of this GI pathogen (Anon 2016). However, a number of other pathways/sources are likely to play a role including consumption of private water (Anon 2010) and contact with the environment (N. J. Strachan, Gormley et al. 2009).

**Figure 1.1 Incidence and hospitalisation discharge rates of human campylobacteriosis in Scotland**




(N. J. C. Strachan, Rotariu et al. 2013) (Data obtained from Health Protection Scotland (HPS) and Information Services Department (ISD) NHS Scotland).

Risk factors for GI pathogens can denote anything that could be associated with the risk of disease (Giesecke 2002). They can be categorized into either source or population attributable risk factors (MacRitchie, Hunter et al. 2013). Source risk factors are directly associated with the pathway of infection; for example, the environment (e.g. contact with farm animals (Howie, Mukerjee et al. 2003)), water exposure (drinking from private water supplies (Anon 2010)) and food exposure (consumption of contaminated chicken meat (Gormley, Macrae et al. 2008)). Population attributable risk factors have an indirect association with infection and include age (N. J. C. Strachan, Rotariu et al. 2013), population density (Ethelberg, Simonsen et al. 2005), and deprivation (Simonsen, Frisch et al. 2008).

## 1.1.1 Deprivation and campylobacteriosis

Deprivation can be defined in several different ways. In Scotland, the Carstairs Score is a measure of access to "those goods and services, resources and amenities and of a physical environment which are customary in society" (McLoone 2004). It is a socio-economical index for the Scottish population that was derived by combining several variables (e.g. number of cars owned per household, male unemployment, overcrowding etc.) to generate indices at postcode sector level. The last available update was based on the 2001 census. The Scottish Index of Multiple Deprivation (SIMD) defines deprivation "as the range of problems that arise due to lack of resources or opportunities covering health, safety, education, employment, housing and access to services as well as financial aspects" (Anon 2012). The SIMD is based on datazones (comprising on average 800 people) of which there are approximately 6,500 in Scotland. The SIMD is based on 7 domains (employment, income, health,

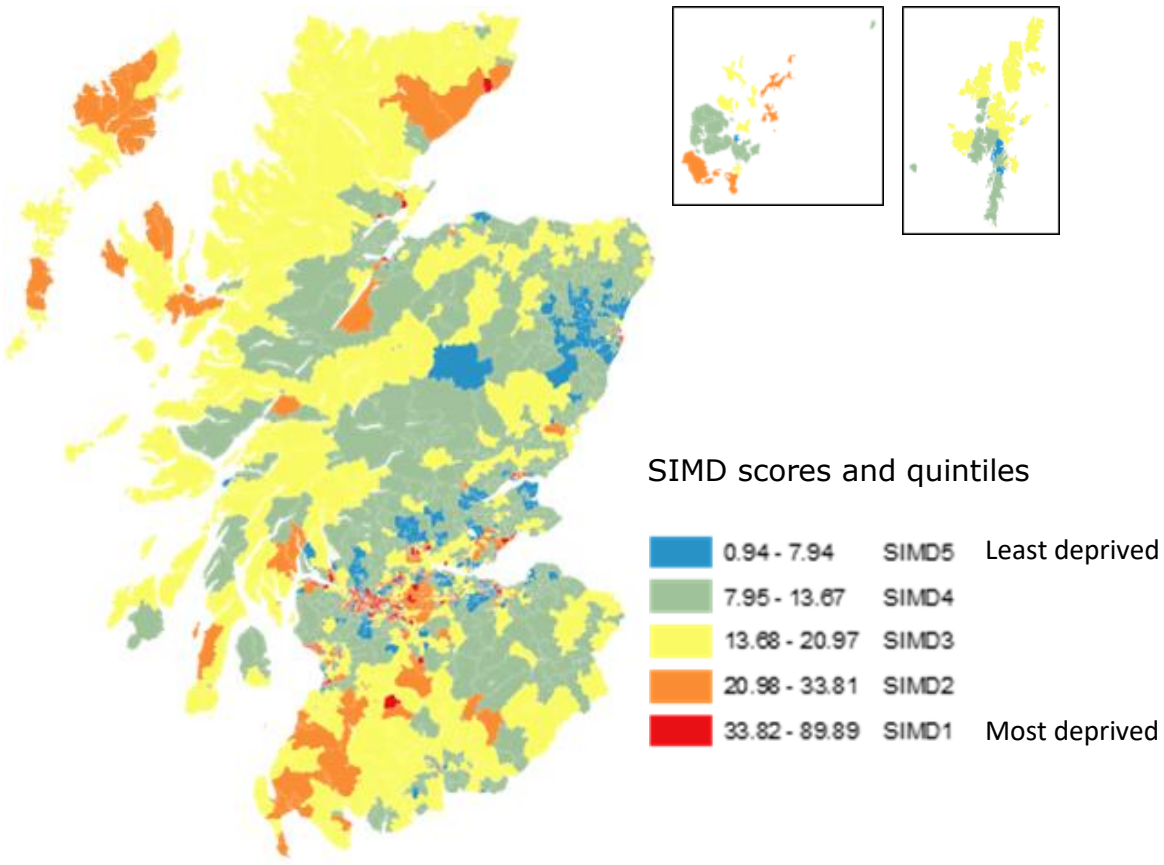
education/skills/training, geographic access to services, crime and housing) constructed from 38 indicators. It measures deprivation, not affluence: datazones with lower scores are less deprived (i.e. contain fewer deprived people). SIMD 2012 data have been used in this project as 2016 data only become available towards the end of the study.

The SIMD scores can be grouped into 5 quintiles, each comprising 20% of the Scottish population, where quintile 1 comprises datazones where the *overall* level of deprivation experienced by residents is highest and quintile 5 where it is lowest. In this report, the short-hand term 'increasing affluence' will be used to describe the transition from more deprived areas to less deprived ones. Careful consideration should be taken in terms of the meaning of affluence in this context. For example a more affluent area does not necessarily mean that there are proportionally more rich people living there, rather that, *overall*, residents are relatively less deprived.

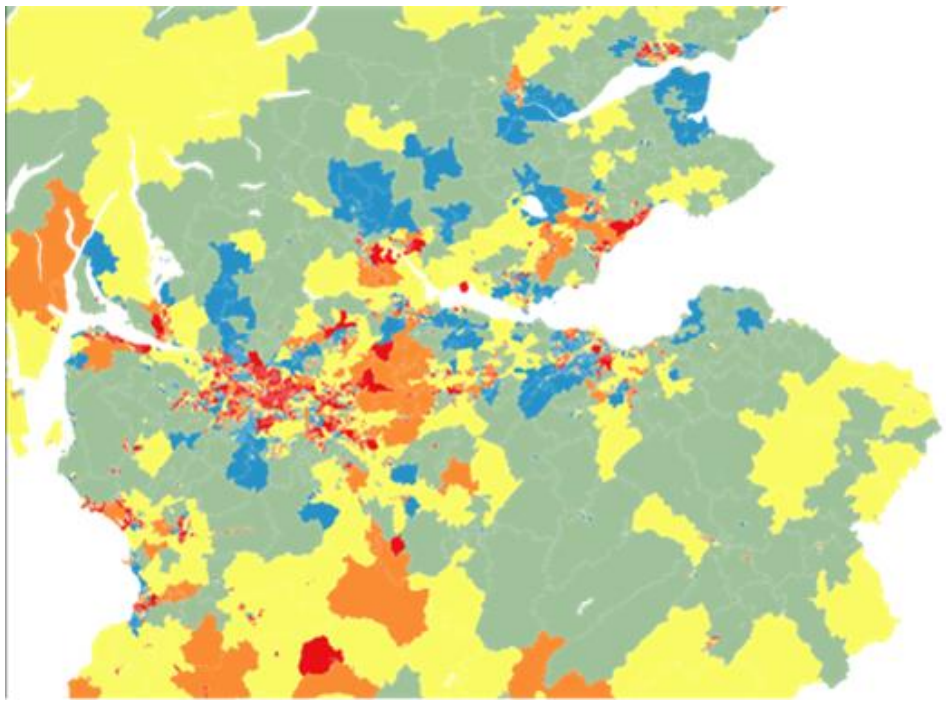
Figure 1.2 provides a map of the SIMD split into 5 quintiles for each data zone across Scotland.

**Figure 1.2. Distribution of SIMD2012 scores and quintiles across (a) Scotland and (b) the central belt.**

(a)




(b)



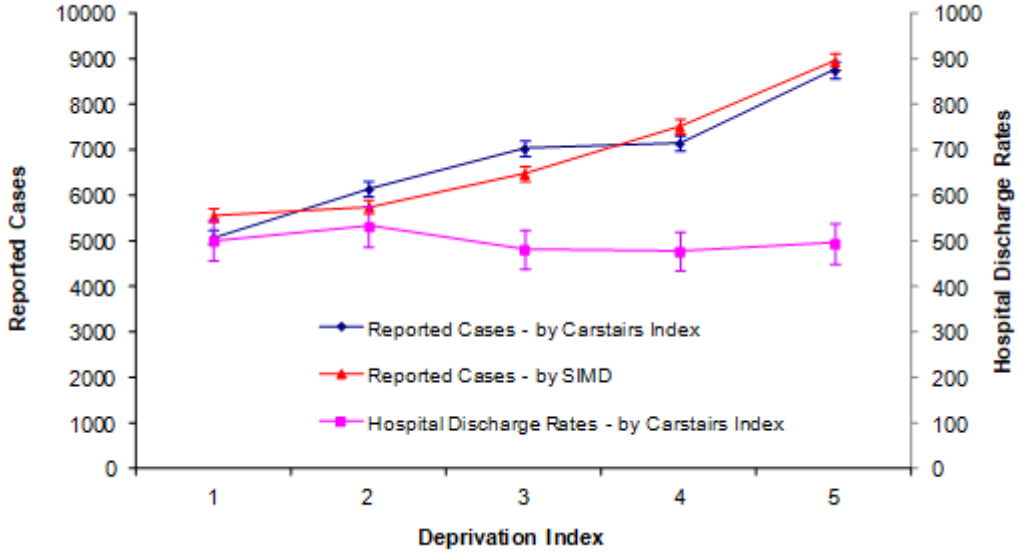
There is growing evidence that the population attributed risk factor, deprivation, is protective for *Campylobacter*. For example studies of reported cases in Scotland (Bessell, Matthews et al. 2010), England & Wales (G. L. Nichols, Richardson et al. 2012), New Zealand (Spencer, Marshall et al. 2012, Sears 2009) and Denmark (Simonsen, Frisch et al. 2008) all suggest that deprivation is protective. Indeed, during the period 2000-2006 there was an excess of 8,700 (26% of all cases) in the four least deprived quintiles of the Scottish population (unpublished data). There are a number of putative explanations for this phenomenon and these include:

1. Differences in culinary habits. Are individuals living in more deprived areas of Scotland more likely to prepare/consume processed or frozen rather than fresh meat, and less likely to eat out in restaurants?
2. Differences in levels of environmental exposure. Are individuals living in more deprived areas of Scotland less likely to be exposed to environmental risk factors (e.g. cattle and sheep faeces, private water supplies etc.) than those living in more affluent areas due to differences in leisure activities and/or access to the countryside?
3. Differences in disease severity. Are individuals living in more deprived areas of Scotland being exposed to less pathogenic strains of *Campylobacter* than those living in less deprived areas? Are there any differences in disease severity in these groups which could explain differences in exposure?
4. Differences in reporting. Are individuals living in more deprived areas of Scotland less likely to seek medical attention for gastrointestinal illness? Are there any differences in the numbers of faecal specimens taken by GPs in some areas of Scotland compared with others?
5. Difference due to foreign travel. Are individuals living in more deprived areas less likely to travel abroad to parts of the world where the risk of campylobacteriosis is high?

A cross-sectional population survey in Grampian (MacRitchie, Hunter et al. 2013) showed that those living in affluent areas had greater exposure to *Campylobacter* source risk factors (e.g. visiting farms and crossing fields, handling farm animals, contact with live chickens, contact with fresh/salt water, and use of a private water supply). Also, in New Zealand that whilst, as already stated, deprivation is protective in urban areas it was found that in rural areas reported rates were not associated with social deprivation index (Spencer, Marshall et al. 2012).

Contrary to the above evidence that deprivation is protective in terms of reported campylobacteriosis cases there is a growing body of evidence that this is not the case for hospitalisations. Back in 1999 it was reported in the Lancet (Olowokure, Hawker et al. 1999) that hospital admission rates for gastrointestinal infections were higher in the deprived population. In New Zealand, this was also found to be the case for campylobacteriosis (Sears 2009). Previous unpublished work by the authors suggests that, in Scotland, while hospitalisation rates do not vary, incidence rates of reported cases decrease with deprivation (Fig. 1.3).

**Figure 1.3 Reported cases and hospital admissions of campylobacteriosis in Scotland 2000-2006**




The five deprivation groups (quintiles) ranging from 1 (most deprived) to 5 (least deprived) each comprise approximately 1 million individuals. Error bars denote 95% bootstrapped confidence intervals. (Note how similar the reported case results are for the Carstairs and SIMD scores).

## 1.1.2 Analytical Epidemiological Methods

A number of analytical methods have been employed to identify risk factors and/or putative sources of human campylobacteriosis.

*Case-control Studies:* The case-control study is an analytical epidemiological method that compares risk factors of people who have been ill (e.g. with campylobacteriosis) with a control group who have not been ill (Giesecke 2002). A meta-analysis (Domingues, Pires et al. 2012) of case-control studies on campylobacteriosis from across the world found that international travel, followed by consumption of undercooked chicken, environmental exposure (drinking water, recreational water use, contact with bird droppings) and direct contact with farm animals (particularly associated with young children) and pets were significant risk factors. Other important factors included pre-existing chronic disease, eating chicken in a restaurant, eating poultry and consuming unpasteurized dairy products.

A case-control study in North-East Scotland (Anon 2010) reported that proton pump inhibitors (PPIs) (Odds ratio (OR) 2.4), overnight stay outside study area (OR 2.03), contact with farm animals (OR 1.50), pets at home (OR 1.23), private water supply (OR 2.98), barbeque and picnic (OR 1.47) and diving in the sea (OR 4.14) were associated with disease whilst consumption of pre-packed ready to eat foods was protective (OR 0.60).

Case-control studies therefore indicate that an overnight stay outside the study area and foreign travel are risk factors for campylobacteriosis. In NE Scotland it was found that 17% and 18% of cases were associated with travel abroad and travel out with the study area (Strachan et al., 2013b). An FSAS study (S14004) indicated that

deprivation was still a protective factor when foreign travel associated cases were excluded. An explanation of the trend that deprivation is protective for both foreign travel and non-travel cases is unknown and is worth investigating further.

*Case-case Studies:* Case-case methodologies have been used when trying to identify risk factors between two different pathogens (or pathogen types). For example a case-case study demonstrated that *C. coli* cases were more likely to drink bottled water, eat pâté, and on average be older than *C. jejuni* cases (Gillespie, O'Brien et al. 2002). This can also be used to determine whether the importance of risk factors has changed over time. Case-case studies remove the differential recall bias that occurs in case-control studies (McCarthy, Giesecke 1999). However, a problem with case-case studies is that those risk factors that are common to the two groups will not be identified. For example, if eating undercooked chicken is of similar importance for contracting human campylobacteriosis in both groups then it will not be seen as a risk factor in the case-case study. There is the potential to use the case-case methodology to compare cases from deprived and non-deprived populations. This technique is likely to identify risk factors that vary between these groups.

### 1.1.3 The surveillance reporting pyramid

Only a fraction of community cases with infectious intestinal disease (IID) are actually reported. For campylobacteriosis this is estimated to be 1 in 9 (C. Tam, Viviani et al. 2011). Reporting of campylobacteriosis can be represented by a reporting pyramid with cases, whether symptomatic or not, in the community located at the bottom and those finally reported by surveillance system at the top (see for example Figure 3.1). Each step in the pyramid (e.g. from community, to GP, to diagnostic lab and finally reporting to the national surveillance system (ECOSS)) offers an opportunity for cases to be omitted from the reporting process.

A UK wide telephone survey to determine rates of diarrhoea and vomiting in the public and also a GP presentation study found that rates of infectious intestinal disease did not vary by deprivation (C. Tam, Viviani et al. 2011). However, it should be noted that *Campylobacter* causes only a small fraction of total IID and that Scotland only comprises a small percentage (8.3%) of the total population of the UK.

A previous FSAS study (S14004) indicated that all diarrhoeal stool samples submitted to clinical labs were tested for *Campylobacter* and that differences in the microbiological methods between the reporting laboratories did not explain the differences between NHS board reporting rates. To estimate under-ascertainment of campylobacteriosis by looking at each level of the pyramid was not feasible in the current study (i.e. not financially possible to sample all cases of diarrhoea in a specified community for *Campylobacter* and determine under ascertainment in reporting with sufficient statistical power at a reasonable cost). However, there was a need to better understand whether there are likely to be any reporting biases, particularly at the community and GP levels in Scotland with regard to deprivation and campylobacteriosis.

It is potentially possible to gather data at the community level on whether individuals are likely to present to a GP if they have a gastrointestinal illness. This can be achieved by asking controls (from a case-control study) this question and then seeing

if the response is different between deprived and non-deprived respondents. This approach does assume that individuals who do not attend a GP when ill are not part of the same group who decline to complete a case-control questionnaire.

At the GP level it is possible to identify whether there is any bias in reporting between deprived and non-deprived cases of GI infection by asking GPs about how they deal with these cases (whether they request a stool sample or otherwise). Further, identifying whether there are any differences in GP behaviour between predominantly deprived or affluent areas would also enable detection of biases in reporting at this level.

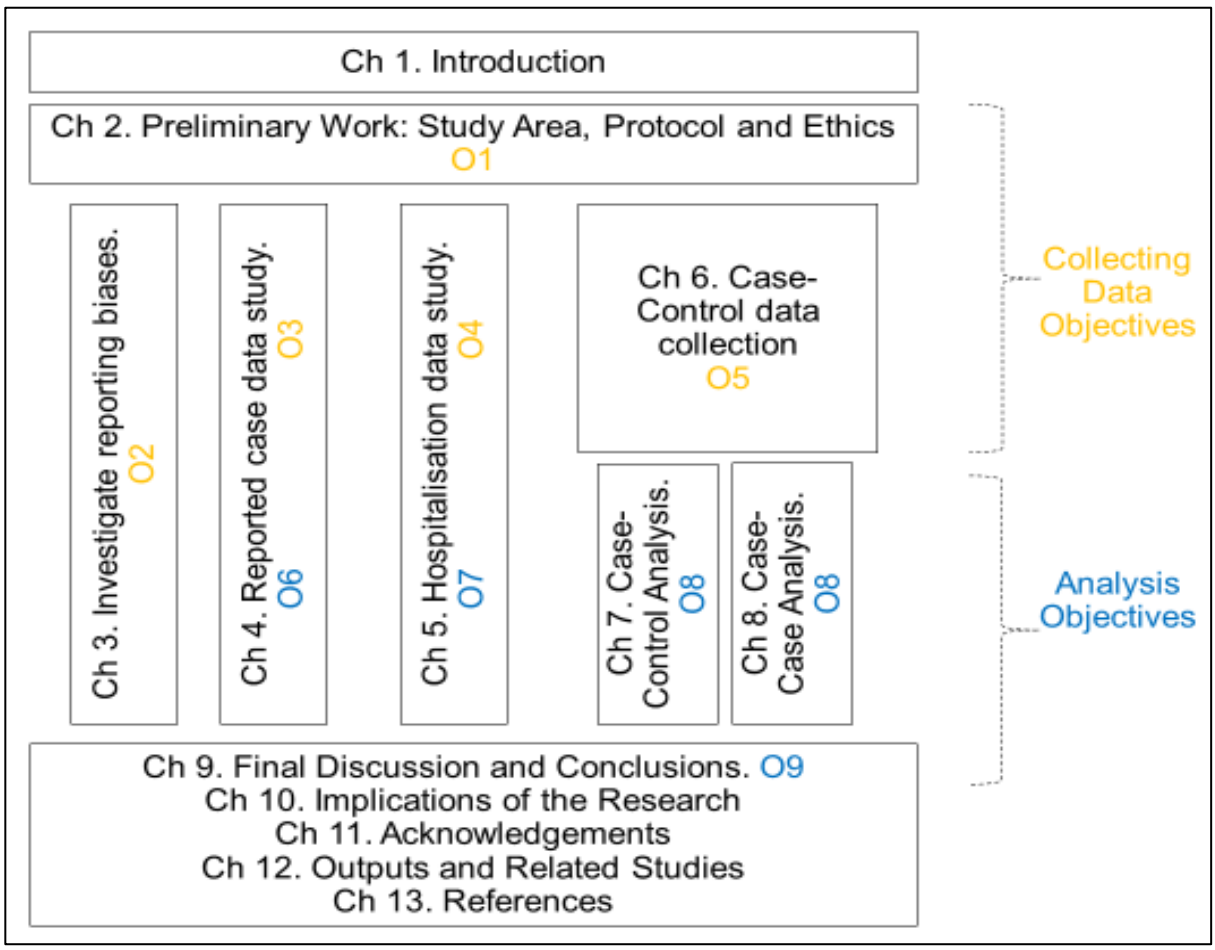
The Practice Team Information (PTI) study run by the Information Services Division (ISD) of the NHS records GP diagnoses from approximately 60 practices across Scotland up until the last part of 2013. In total >10 million records were stored that detail the illness/diagnoses of the patient. These diagnoses included 'Gastroenteritis of presumed infectious origin' (RCG3 code). However, the practices did not collect information regarding submission of stool samples but these data are collated by the NHS medical microbiology diagnostic laboratories (MMDLs). Linking these data together enables the determination of the number of stools submitted to the MMDLs as a fraction of the population of each practice diagnosed with a gastrointestinal infection (RCG3). There is the potential to stratify these results by deprivation (e.g. SIMD) to identify if there are any differences between those practices that serve populations that are more or less deprived.

## 1.2 Objectives

This project had four phases. The first was the development of protocols and obtaining ethical approvals, the second was collection of data, the third analysis of data and the fourth the writing up of results and submission of reports to FSS. This was broken down into the following nine objectives (Figure 1.4).

- Objective 1 Finalise study area, prepare protocol, questionnaire and ethics application (Chapter 2)
- Objective 2 Investigate potential reporting biases by level of deprivation (Chapter 3)
- Objective 3 Collect reported case data (Chapter 4)
- Objective 4 Collect hospitalisation data (Chapter 5)
- Objective 5 Perform case-control study (Chapter 6)
- Objective 6 Analyse reported case data (Chapter 4)
- Objective 7 Analyse hospitalisation data (Chapter 5)
- Objective 8 Analyse questionnaire data utilising both case-control and case-case formats (Chapters 7 and 8)
- Objective 9 Prepare and submit annual and final report (this document)

**Figure 1.4. Overview of Study and Objectives**



More specifically this report investigates:

- Campylobacteriosis reporting biases
    - at the community level
    - at the GP level
    - at the reported case level
- Reported cases
    - whether deprivation is protective
    - whether rurality/urbanicity and deprivation are linked
    - whether adjacency to a GP practice and deprivation are linked
- Hospitalised cases
    - whether deprivation is protective
    - whether rurality/urbanicity and deprivation are linked
    - whether adjacency to a hospital and deprivation are linked
- Case-Controls and Case-Cases
    - to identify which factors are a risk and which are protective for campylobacteriosis
    - to determine whether deprivation plays a protective role for campylobacteriosis
    - to determine which risk factors are different for the most and least deprived

# 2. Study Area, Protocol, Questionnaires and Ethics

## 2.1 Introduction

Ideally, the study would include the whole Scottish population in order to maximise the rate of data accrual and optimise statistical power. However, the most and least deprived populations are heterogeneously distributed across Scotland. Greater Glasgow and Lanarkshire NHS boards have the largest deprived populations whereas Grampian and Lothian have the most affluent. NHS Public Health teams, where available, were invited to collaborate based on these considerations. Participating NHS boards are detailed in Table 2.3.

The study also required the design of case-control questionnaires that would enable identification of risk factors for human campylobacteriosis with an emphasis on differentiating between the least and most deprived populations. The questionnaire design aimed to help identify potential biases in reporting between the different populations (See Chapter 3 for further details).

Since the study involved sending questionnaires to patients with a clinical case of human campylobacteriosis, it was necessary to obtain ethical approval from both the Research Ethics Committee (REC) and the NHS Research and Development offices (NHS/HSC R&D). Approval from the Public Benefit and Privacy Panel (PBPP) was also necessary because the study used NHS data and, in particular, required access to the Community Health Index (CHI) database primarily for identification of controls.

## 2.2 Finalise study area

### 2.2.1 Statistical Power

#### 2.2.1.1. Method

The statistical power is the likelihood that a study will detect an *effect* (e.g. an outcome, a result, a difference in exposure to a risk factor between two population groups) when there is an *effect* to be detected (https://effectsizefaq.com/2010/05/31/what-is-statistical-power/). This study looked for differences between cases of human campylobacteriosis and controls in terms of exposure to risk factors (e.g. eating chicken outside home, contact with animals etc.). This study also looked for similar differences between populations from least and most deprived populations using a case-case approach.

In this study odds ratios (ORs) are used to quantify the differences between cases and controls for specific risk factors. The statistical power in this case is the likelihood (%) to detect a minimum odds ratio of 2.00 between cases and controls with 95% confidence (i.e. confidence level - $\alpha = 0.05$) assuming that 4% of controls are exposed to the risk factor (Efird 2013).

If the proportion of exposed people in the control group is $p^+_{controls}$, and OR is the odds ratio between cases and controls, then the proportion of exposed cases will be

$$p^+_{cases} = \frac{OR\ p^+_{controls}}{1+(OR-1)p^+_{controls}} \quad (2.1)$$

An algorithm was developed using the @Risk (http://www.palisade.com) add-in for Excel to determine the statistical power. Briefly, the number of questionnaires obtained from cases and controls were $N_{\text{cases}}$ and $N_{\text{controls}}$ respectively. The probability of control and cases being exposed to a risk factor is $p^+_{controls}$ and $p^+_{cases}$ respectively.

A Monte Carlo simulation was conducted with 10,000 iterations.

It is assumed that the number of exposed controls follows a binomial distribution and sampled as follows:

$$N^{exp}_{controls} = Binomial(\text{N}_{\text{controls}}, p^+_{controls}) \quad (2.2)$$

Hence the number of controls that are not exposed is:

$$N^{notexp}_{controls} = \text{N}_{\text{controls}} - N^{exp}_{controls} \quad (2.3)$$

Similarly the number of exposed cases also follows a binomial distribution and is sampled as follows:

$$N^{exp}_{cases} = Binomial(\text{N}_{\text{cases}}, p^+_{cases}) \quad (2.4)$$

Hence the number of cases that are not exposed is:

$$N^{notexp}_{cases} = \text{N}_{\text{cases}} - N^{exp}_{cases} \quad (2.5)$$

The simulated odds ratio was calculated as follows

$$OR_1 = \frac{N^{exp}_{cases}/N^{notexp}_{cases}}{N^{exp}_{controls}/N^{notexp}_{controls}} \quad (2.6)$$

and the Fisher's exact test (Fisher 1935) was used to determine if this odds ratio ($i.e. OR_1$) was significantly >1.

This algorithm was repeated 10,000 times, and the percentage of times a significant odds ratio was obtained represents the statistical power.

### 2.2.1.2 Statistical Power Considerations

Statistical power calculations were performed for plausible numbers of completed questionnaires from the study. The target OR was set to 2 and proportion of controls exposed to 4%. Table 2.1 presents simulated data for the statistical power for different scenarios.

**Table 2.1. Statistical power calculations for different numbers of cases and controls.**

| $N_{cases}$ (returns) | $N_{controls}$ (returns) | $p^{+}_{controls}$ (reference) | $p^{+}_{cases}$ | OR | Power (%) |
|---|---|---|---|---|---|
| 1500 | 1500 | 0.04 | 0.077 | 2 | 99 |
| 750 | 750 | 0.04 | 0.077 | 2 | 84 |
| 650 | 650 | 0.04 | 0.077 | 2 | 78 |
| 600 | 600 | 0.04 | 0.077 | 2 | 76 |
| 550 | 550 | 0.04 | 0.077 | 2 | 73 |
| 500 | 500 | 0.04 | 0.077 | 2 | 68 |
| 300 | 300 | 0.04 | 0.077 | 2 | 43 |
| 452 | 500 | 0.04 | 0.077 | 2 | 66 |
| 146 | 52 | 0.04 | 0.077 | 2 | 9 |
| 332 | 113 | 0.04 | 0.077 | 2 | 19 |

Statistical power analysis shows that recruiting 1500 cases and 1500 controls will have 99% power to detect a minimum odds ratio of 2.00 with 95% confidence, whilst for 650 cases and 650 controls the power will be 78% (Table 2.1).

Previous experience, from similar studies, as well as the recent FSAS i-CAMPS-3 project (Contract S14054) suggests that approximately 50% of human *Campylobacter* cases and 25% of controls are likely to return a completed questionnaire. Hence, at the start of the study, the 8 health boards recruited, were expected to provide 7500 cases over two years. In addition for every case there were 2 control questionnaires submitted. At the end of the study there were expected to be approximately 1500 case and 1500 control questionnaires. This would provide a statistical power of 99% from Table 2.1.

With regard to the case-case study since only the cases are used (750 from SIMD1 and 750 from SIMD5) then it would be expected to have a statistical power of 84% (Table 2.1).

The third bottom row of Table 2.1 provides the actual number of questionnaires received in the actual case-control study (Chapter 7) for domestically acquired cases and controls (452 cases and 500 controls respectively). The statistical power was 66%.

The second bottom row of Table 2.1 provides the actual number of questionnaires received in the actual case-control study (Chapter 7) for foreign travel associated cases and controls (146 cases and 52 controls respectively). The statistical power was 9%.

The bottom row of Table 2.1 provides the actual number of questionnaires received in the actual case-case study (Chapter 7) for SIMD 5 cases and SIMD1 cases (332 and 153 respectively). The statistical power was 19%.

## 2.2.2 Study Area

### 2.2.2.1 Case control Study

The study area for the case-control study initially consisted of 8 health boards (Fife, Forth Valley, Greater Glasgow & Clyde, Grampian, Highland, Lanarkshire, Lothian and Tayside). Table 2.2 presents the distribution of the population in these health boards by SIMD quintiles. Although Lanarkshire signed up they were unable to fully participate and so dropped out of the study, therefore the study was left with a total of 7 health boards participating. The case-control study started on 1st June 2016 and the boards joined at various dates (full details in Chapter 3). Questionnaires were submitted until 31st August 2018.

**Table 2.2. The distribution of the Scottish population by SIMD quintiles and health boards.**

| Health Board | Population | SIMD1*(%) | SIMD2(%) | SIMD3(%) | SIMD4(%) | SIMD5(%) |
|---|---|---|---|---|---|---|
| Ayrshire & Arran (AA) | 370,686 | 26.6 | 26.6 | 16.6 | 15.8 | 14.3 |
| Borders (BO) | 114,445 | 4.5 | 13.2 | 32.8 | 42.9 | 6.7 |
| Dumfries & Galloway (DG) | 149,575 | 7.5 | 21.0 | 38.2 | 26.0 | 7.3 |
| Fife (FF) | 369,545 | 18.0 | 20.7 | 19.6 | 20.4 | 21.3 |
| Forth Valley (FV) | 303,672 | 14.3 | 23.6 | 18.9 | 21.8 | 21.4 |
| Grampian (GR) | 586,371 | 5.8 | 11.6 | 21.0 | 26.8 | 34.7 |
| Greater Glasgow & Clyde (GC) | 1,157,517 | 35.1 | 17.9 | 14.5 | 14.0 | 18.6 |
| Highland (HG) | 321,489 | 8.0 | 18.4 | 32.4 | 31.9 | 9.2 |
| Lanarkshire (LN) | 655,911 | 23.5 | 27.7 | 22.1 | 13.7 | 13.0 |
| Lothian (LO) | 875,513 | 11.0 | 19.0 | 18.7 | 19.1 | 32.2 |
| Tayside (TY) | 415,162 | 16.9 | 15.7 | 17.6 | 31.4 | 18.4 |
| Orkney (OR) | 21,804 | 0.0 | 21.3 | 17.2 | 58.5 | 2.9 |
| Shetland (SH) | 23,166 | 0.0 | 3.1 | 39.2 | 50.7 | 7.0 |
| Western Isles (WI) | 27,027 | 0.0 | 36.0 | 61.1 | 2.9 | 0.0 |
| Total | 5,391,883 | 19.0 | 19.5 | 20.1 | 20.8 | 20.6 |

* SIMD1 means most deprived population and SIMD5 least deprived population.

#### 2.2.2.2 Reported case and Hospitalisation Studies

Case data were obtained from the ECOSS database and hospital discharge data were provided by ISD. The study area for these studies comprised the whole of Scotland.

## 2.3 Generation of case and control questionnaires

Questionnaire content (ANNEX 2.1 & 2.2) was informed by previous case-control studies. They consisted of the following sections: general details, household income, details of illness (for cases), likelihood of presenting to a GP following a 'tummy bug' (controls), previous health conditions, travel, exposure to animals, water and food, and additional information. FSS commented on the draft of the questionnaires prior to their implementation. In the case-control study the participants had the opportunity to complete the questionnaires either in paper format or via the web.

## 2.4 Research Protocol

A research protocol for the study was prepared for ethics. The final version (V.4) (Nov 2017) is provided in ANNEX 2.3.

## 2.5 Ethics application

An ethics application (ANNEX 2.4) for the study was prepared and submitted on 2nd July 2015. Approval was received on 18th September 2015 (ANNEX 2.5). On receipt of this a submission was made to PBPP since the study involved access to NHS data on cases and hospitalisations of campylobacteriosis as well as access to the CHI database to obtain details of controls. This request (ANNEX 2.6) was submitted on 7th August 2015 and approval was obtained on 21st December 2015 (ANNEX 2.7). Following completion each health board required a contract to be put in place with the University since there was payment for the work. This could only be progressed once approvals had been given by ethics and PBPP panels. Following this, start-up meetings with each health board team were held and NHS Research and Development approvals were obtained. It took in total 6 months for all of the contracts, start-up meetings and approvals to take place (Table 2.3).

A number of ethical amendments were required during the study (e.g. five for REC and three for PBPP). In the annexes to this Chapter the final versions of the documents are provided.

**Table 2.3. Progress and approvals with Case Control Study.**

| Health Board | Ethics Approval (Date) | PBPP Approval (Date) | Start-up meeting | R&D Approval (Date) | Contract (date) | Date Started Case Control Study |
|---|---|---|---|---|---|---|
| Fife (FF) | Y (18/9/15) | Y(21/12/15) | Y | Y (10/5/16) | Y (16/6/16) | 1/7/16 |
| Forth Valley (FV) | Y (18/9/15) | Y(21/12/15) | Y | Y (25/5/16) | Y (25/5/16) | 7/3/17 |
| Glasgow & Clyde (GC) | Y (18/9/15) | Y(21/12/15) | Y | Y (23/5/16) | Y (13/5/16) | 1/6/16 |
| Grampian (GR) | Y (18/9/15) | Y(21/12/15) | Y | Y (15/5/16) | Y (3/5/16) | 1/6/16 |
| Highland (HG) | Y (18/9/15) | Y(21/12/15) | Y | Y (23/6/16) | Y (21/6/16) | 9/7/16 |
| Lanarkshire (LN) | Y (18/9/15) | Y(21/12/15) | Y | Y (6/7/16) | Y (28/6/16) | 15/8/16 |
| Lothian (LO) | Y (18/9/15) | Y(21/12/15) | Y | Y (26/4/16) | Y (16/6/16) | 15/8/16 |
| Tayside (TY) | Y (18/9/15) | Y(21/12/15) | Y | Y (19/09/17) | Y (13/5/15) | 15/6/17 |

## 2.6 Overview of Ethics Process

The project started on 5th January 2015 but questionnaires were not submitted to cases and controls until June 2016 (a total of 17 months). It had been envisaged that a period of 5 months would be sufficient to achieve approvals for this study (3 months to submit and 2 to obtain the approval). However, obtaining these approvals took considerably longer.

Preparation for the REC submission took longer than anticipated (5 months). The reason for this was that additional time was required for the protocol, questionnaires and REC forms to be reviewed by University Research Governance as this needed to be done prior to submission to the REC. Also, the overall process was complex because of the multiple studies that were included in the project. Once submitted to the REC, approval took 2.5 months.

Submission to PBPP occurred in August 2015. This was a complex task because the PBPP panel had just been set up and so this project was the first to go through the process. PBPP approval was given at the end of December 2015. Following this it took approximately six months to obtain R&D approvals and contracts and hence the study started in June 2016.

Although arriving at the point of sending out questionnaires took 17 months and considerably more effort than estimated in the project, which the University and individual staff had to bear, it is likely that this will be a smoother process in future. However, because there are so many steps and individuals involved, in the authors' view, this would be difficult to achieve in much less than one year. This is worth bearing in mind for any future multi-site case-control studies.

All of the health boards in Table 2.3 were involved throughout the study except Tayside, Forth Valley and Lanarkshire. Tayside started one year and Forth Valley nine months into the project when staff there became available. Lanarkshire started sending questionnaires at the start of the project but had virtually zero returns. It was unclear why this was the case but as a consequence Lanarkshire was withdrawn from the case-control study.

## 2.7 Conclusion

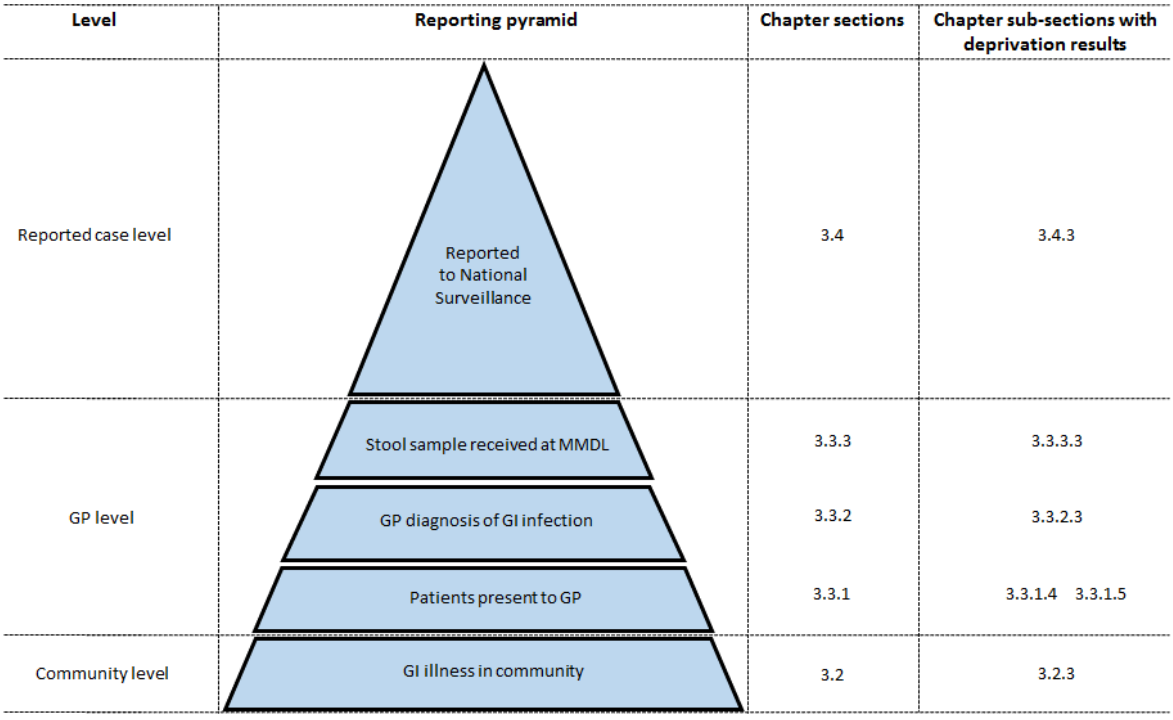
All of the permissions required to carry out the study were obtained but this took approximately 17 months. Seven health boards participated comprising 87% of the Scottish population and statistical power calculations were carried out on this basis. For the actual returns the statistical power was 66% for the domestic case-control study, 9% for the foreign travel associated case-control study and 19% for the case-case study. This is based on an odds ratio of 2.0 with 4% exposure in the control population.

# 3. Potential reporting biases by level of deprivation

## 3.1 Introduction

Differences in the incidence and/or hospitalisation rates of human campylobacteriosis cases between the least and most deprived populations in Scotland could be due to differences in reporting. This may occur at one or more levels of the surveillance pyramid (Figure 3.1). This chapter utilises data from the case-control study, hospital discharges (as a proxy for hospitalisations), reported cases, GP questionnaires, Medical Microbiological Diagnostic Laboratories (MMDL) returns and Practice Team Information (PTI) diagnoses of gastroenteritis of infectious origin to establish whether any biases were detectable.

**Figure 3.1 The reporting pyramid and chapter structure.**




Specifically, this chapter considers bias at three levels, community (Section 3.2), GP (Section 3.3) and reported case (3.4). Ideally, this would include the study of the rate of campylobacteriosis at each level of the pyramid. However at the community and GP levels, information is only available in terms of Infectious Intestinal Disease (IID) and this is what is examined. In particular, the analysis at the community level looks at the likelihood of individuals with an IID presenting to a GP. Whilst the analysis at the GP level involves three different

studies: interviews and questionnaires of GPs (3.3.1); Practice Team Information on GP diagnoses of infectious intestinal disease (3.3.2) and microbiological reporting of cases by MMDLs (3.3.3).

In addition, campylobacteriosis symptoms are compared, at the reported case level (3.4)

(1) by deprivation (i.e. SIMD1 with SIMD5) and
(2) by intensity of health care (hospitalised versus not hospitalised cases).

Finally, underreporting between the different steps of the pyramid is collated together and discussed in section 3.5.

## 3.2 Estimation of reporting biases at the community level

### 3.2.1 Aim

This section investigates reporting rates from community GI illness to likelihood of presentation to a GP (see Figure 3.1).

### 3.2.2 Data and methods

Estimation of reporting bias at the community level is based on questionnaire responses from the control group (from SIMD1 and SIMD5) in the case-control study (see Chapter 6).

Control subjects were asked about (i) their views on when and why they would consult a GP if they had a case of gastroenteritis of infectious origin and (ii) if they would supply a stool sample if asked by the doctor/nurse (see Annex 2.2 Control Questionnaire, Section B, p18).

The proportion of the population who will make a doctor's appointment was calculated based on the duration (up to 14 days) and type of symptoms (e.g. diarrhoea or loose stools, nausea or vomiting, abdominal pain / stomach cramps, blood in stools and fever). Figures were generated to illustrate these data for the SIMD1 and SIMD5 populations combined, as well as the most (SIMD1) and least (SIMD5) deprived populations separately.

Further quantification of any potential biases was performed by univariate logistic regression (Cox 1958, Kleinbaum, Klein 2010), which enabled odds ratios (ORs) and statistical significance (P-values) to be calculated.

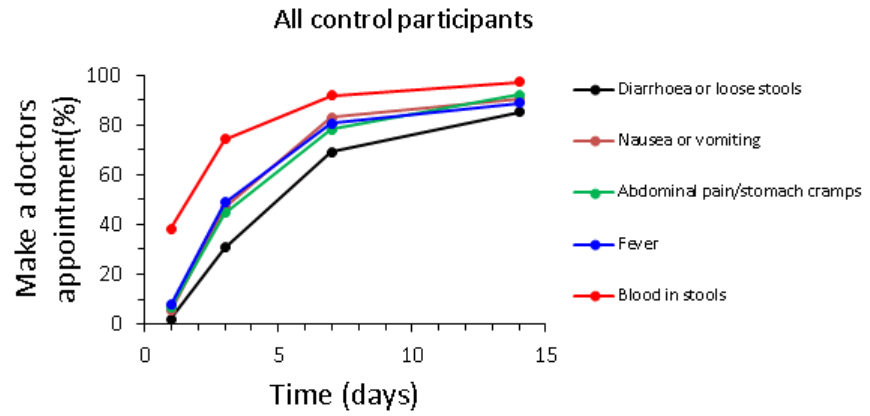
To determine the under-reporting rate two extremes were used. The first was the proportion of the population making a doctor's appointment with the mildest symptom (diarrhoea or loose stools) and shortest duration (1 day). The second, repeating the calculation but with any symptom that lasts for the longest duration (14 days).

### 3.2.3 Results and discussion

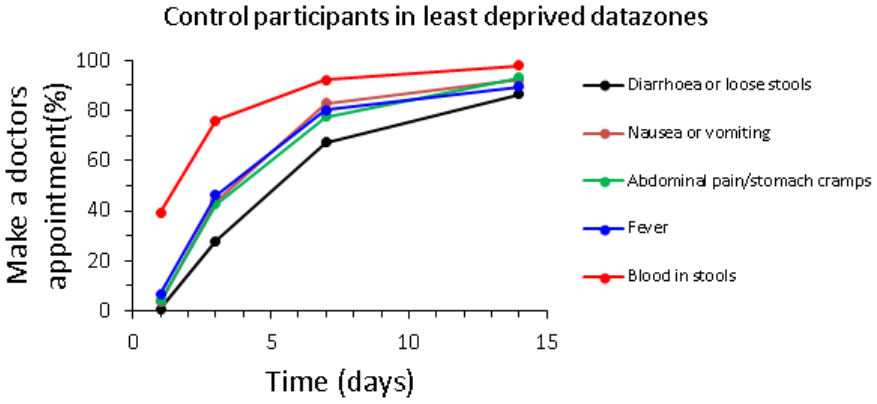
Figure 3.2(a) illustrates that people are generally more likely to make a GP appointment if they have blood in their stools and least likely if symptoms are diarrhoea or loose stools for any given symptom duration. This is as anticipated: features (such as bloody stools) that are perceived as more severe or unusual would be expected to prompt a higher rate of GP consultation. The same general pattern is observed when stratified by deprivation (Figures 3.2(b) and (c)).

**Figure 3.2 Percentage of control participants responding that they will make a doctor's appointment after experiencing a particular symptom for a number of days**

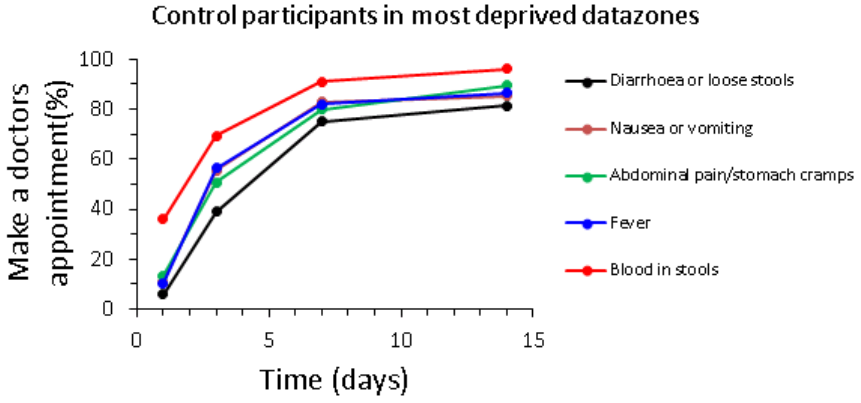
(a)




(b)



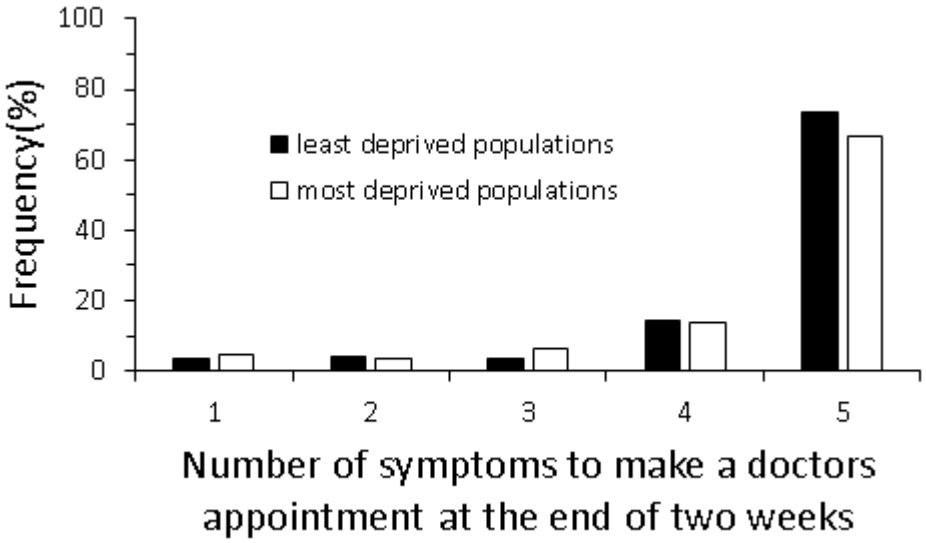

(c)




(a) All controls (n=552), (b) controls living in SIMD1 data zones (population quintile comprising largest number of deprived people, n=139) and (c) controls living in SIMD5 data zones (population quintile comprising smallest number of deprived people, n=407). Note: Analysis excludes 6 control patients who did not have SIMD information available.

The number of symptoms required to make a doctor's appointment at the end of two weeks is presented in Figure 3.3. Any number of symptoms is found by adding up the black bars for the least deprived population (98.8%) and white bars for the most deprived population (95.0%). It was found that the number of symptoms at 14 days that would prompt a GP appointment did not significantly differ ( $\chi^2$(4 degrees of freedom, N=534)=2.83, P=0.59)) by deprivation (Figure 3.3). However, an apparent difference is found for nausea or vomiting of 14 days duration (Table 3.1). Here, respondents from the least deprived population are more likely to make a GP appointment.

**Figure 3.3 The distribution of control respondents by number of symptoms**




The *distribution* of control respondents by number of symptoms required to make a doctor's appointment after 14 days - stratified by deprivation. The symptoms were: diarrhoea or loose stools, nausea or vomiting, abdominal pain/stomach cramps, fever and blood in stools.

**Table 3.1 Likelihood of making a doctor's appointment after falling ill with a gastrointestinal infection after 14 days of a particular symptom.**

| | Least deprived* | | Most deprived* | | | |
|---|---|---|---|---|---|---|
| Symptom | Make appointment (%)*** | Not make appointment (%)*** | Make appointment (%)*** | Not make appointment (%)*** | OR (95%CI)** | P-value |
| Diarrhoea or loose stools | 344(86.4) | 54(13.6) | 108(81.2) | 25(18.8) | 0.68(0.40,1.14) | 0.159 |
| Nausea or vomiting | 368(92.2) | 31(7.8) | 115(85.2) | 20(14.8) | 0.48(0.27,0.88) | 0.026 |
| Abdominal pain / stomach cramps | 368(93.2) | 27(6.8) | 120(89.6) | 14(10.4) | 0.63(0.32,1.24) | 0.192 |
| Blood in stools | 391(97.8) | 9(2.3) | 126(96.2) | 5(3.8) | 0.58(0.19,1.76) | 0.349 |
| Fever | 351(89.5) | 41(10.5) | 115(86.5) | 18(13.5) | 0.75(0.41,1.35) | 0.342 |

*The total number of individuals in each deprivation group who answered question 3.1 of the control questionnaire that they would/would not make a doctor's appointment following 14 days duration of the particular symptom.
** Odds ratio calculated as: (Most deprived "Yes"/ Most deprived "No")/(Least deprived "Yes"/ Least deprived "No")
***Some of the respondents did not complete all of the relevant sections of the questionnaire and as such the totals for make an appointment and not make appointment for each symptom varies.

The data from the claimed self-reporting of symptoms has been used to calculate any under-reporting from the community reporting pyramid. To do this, two extremes of a severe symptom lasting 14 days and a mild symptom of 1 day duration was used. This ranged from 1.0 (= 531/517 i.e. 98% of individuals will consult a doctor) for the most severe symptom (blood in stools) lasting up to 14 days (data from table 3.1) to 48 (=531/11 i.e. 2.1% of individuals will consult a doctor) for the mildest symptom diarrhoea of 1 day duration (data underlying Figure 3.2(a)).

## 3.3 Estimation of reporting biases at the GP level

The assessment of GP level reporting bias is approached in three ways: GP interviews and questionnaire, PTI study and MMDL study. Each of these is detailed in turn in the following sections.

### 3.3.1 GP interview and questionnaire study

#### 3.3.1.1 Aim(s)

The aim of this study was to develop insight into potential biases at the point of GP consultation that might result in some groups of patients, all else being equal, being more likely to be invited to submit a stool sample. The outputs from initial semi-structured interviews with a small group of GPs was used to inform the design of a larger GP questionnaire to explore and evaluate potential biases.

#### 3.3.1.2 GP interview methods

Interviewees were purposively selected to include a diversity of GP/individual (age, sex, duration of practice) and practice setting (rurality and socio-economic background of patient-list) characteristics. In particular, GPs from relatively affluent or deprived practices were chosen as well as those who saw a mixed range of patients with respect to this characteristic. This strategy aimed to optimise the range of experiences and thus likelihood of identifying potential biasing factors. GP (and Practice) characteristics are presented in Table 3.2. GPs with concurrent experience of both relatively deprived and affluent patients were anticipated to be most valuable for the elicitation of potential significant factors. The interviews were semi-structured, with a list of questions/points that were pre-prepared to guide the interview process (Annex 3.1).

**Table 3.2 GP interviews Characteristics & settings.**

| Interview Date | Age | Sex | Setting | Patient affluence | Years as GP |
|---|---|---|---|---|---|
| 7/2/17 | 49 | F | Urban/rural | Mixed | 21 |
| 22/3/17 | 47 | F | Post-industrial urban | Deprived (Mixed out of hours) | 8 |
| 11/4/17 | 47 | M | Rural | Affluent | 13 |
| 12/4/17 | 40 | F | Post-industrial | Deprived | 11 |

### 3.3.1.3 Process and Use of GP interview information

The notes taken at the time of the interview were typed up later on the interview day and the four sets of responses to interview carried out are summarised in Annex 3.1.

Information was gathered on GP perceptions of the stool sampling process from patient presentation through sample request to submission and feedback of results. The two crucial steps in terms of bias introduction were identified as:

1) decision of the health care professional (historically, usually GP) to request a sample and

2) compliance of the patient to then provide and submit a sample.

A list of factors that may play a role in these decisions (such as travel, severe diarrhoea and employment) were collated from the GP interviews and were included in the questionnaire (Annex 3.2).

### 3.3.1.4 GP Questionnaire design and strategy

Questionnaire design: An anonymised GP questionnaire (Annex 3.2) was designed based on the responses from the semi-structured interviews (n=4) (see 3.3.1.2 above). These interviews were used to inform the format of two questions to be asked: (i) how a GP decides whether or not to take a stool sample and (ii) what in the GP's opinion influences the likelihood of stool sample submission by patients. The questionnaire also gathered information on the GP's age, gender, year started work as a GP and year started to work in the current GP practice. To refine this study by the deprivation status of the population registered in each practice, the questionnaires were marked as "1" or "5" if a large proportion (>70%) lived in data zones classified as SIMD1 or SIMD5 respectively. The questionnaire had "Likert-type" (Croasmun, Ostrom 2011) answer choices on a series of risk factors including disease symptoms, travel abroad, socio-economic status, existence of similar cases, etc. (see questionnaire in Annex 3.2).

The questionnaire asked 3 questions:

Question 1 in the questionnaire asked about personal details of the GP (age, sex, year started work as GP and year started work in current practice).

Question 2 asked the GPs "Consider a patient who has presented to you with possible infectious gastroenteritis. How much would each of the following factors increase the likelihood that you request a stool sample?" There were a number of factors that that were listed and the GP could respond to one of five options that ranged from "Very important" to "Not Relevant". The response for each factor was calculated as an appropriate percentage. This was done for all GPs and those serving predominantly SIMD1 or SIMD5 patient. A Mann-Whitney test was done for each factor comparing SIMD1 and SIMD5 to see if the responses were significantly different (Mann, Whitney 1947).

Question 3 asks GPs " For such a patient that you have asked to submit a stool sample, please rank the 5 factors that you think would most influence them to actually submit a stool sample." The GPs were given the same list of factors as in question 2. They were then asked to rank the top 5 factors with 1 being most important. Percentage responses were calculated, heat maps generated and Mann Whitney test performed between SIMD1 and SIMD5 as described above.

Questionnaire strategy: A list of GP practices (n=950 in 2018) in Scotland were downloaded from the ISD website (https://www.isdscotland.org/Health-Topics/General-Practice/Publications/2016-12-13/2016-12-13-GPWorkforce2016-Report.pdf?321596861). In addition the population (n=5,652,871) that these practices serve stratified by SIMD deprivation quintiles was obtained from ISD.

It was observed that in 2018, forty-two GP practices have more than 70% of their registered population (totalling 127,400 individuals) in the most deprived quintile (SIMD1) and forty-one of these are in the study area of this project (37 in Greater Glasgow & Clydeside, 3 in Lothian and 1 in Fife. Also, thirty-one GP practices have more than 70% of their registered population (180,000 individuals) in the least deprived quintile (SIMD5), all of which are in the study area of this project (13 in LO, 12 in GG&C and 6 in GR). For this sub-study questionnaires were submitted by post to a shortlist of GPs in two rounds. The first round sent questionnaires to one GP in each practice (41 in most deprived and 31 in least deprived). In the second round, questionnaires were sent to the GP practices, but to a different named GP, except for those where there was only one GP. In total, 136 letters were sent of which 76 and 60 were to most deprived (SIMD1) and least deprived (SIMD5) practices respectively.

#### 3.3.1.5 Results and discussion from GP questionnaires

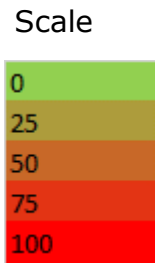
Fifty-six questionnaires (41%) were returned by GPs, 29 (38%) from practices located in SIMD1 regions and 27(45%) from SIMD5 ones. Forty-four percent of the GPs respondents were male and 56% female. The average age of the GP participants was 51.4 years (range 33 to 69 years). They had worked as a GP for an average of 24.5 years (range 6 to 41 years); and in the current GP practice

for 19.4 years (range 1 to 39 years)). No difference (95 percentiles overlapped) in terms of age, gender and work experience as GP, could be found between those working in SIMD1 compared to SIMD5 regions.

Sixty-six percent of the GPs considered "recent foreign travel" (66.1%) as a very important factor to request a stool sample from patients. This is followed by "prolonged diarrhoea" (53.6%) and "existence of similar cases" (51.8%) (Figure 3.4). Further, 96% of the GPs considered "prolonged diarrhoea" as "important" and "very important" when they decide to request a stool sample. Socio-economic factors (e.g. "patient living in an affluent area" or "unemployment") are considered not relevant by most GPs when deciding whether to ask patients for a stool sample. This ranking is confirmed by the Friedman signed-rank test ($P<0.001$). The heat map patterns in Figure 3.4 also illustrate the lack of difference by deprivation in the GPs' responses about stool sample submission. This is consistent with the Mann-Whitney's non parametric test ($P>0.05$ when comparing risk factors between both deprivation categories).

## Figure 3.4 Heat map showing the importance of risk factors in GPs decisions on requesting stool samples

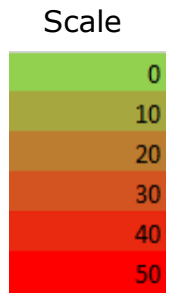
| | All GPS | | | | | GPs in most deprived practices | | | | | GPs in least deprived practices | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2. Consider a patient who has presented to you with possible infectious gastroenteritis. How much would each of the following factors increase the likelihood that you request a stool sample? | Very Important (%) | Important (%) | Useful (%) | Sometimes useful (%) | Not relevant (%) | Very Important (%) | Important (%) | Useful (%) | Sometimes useful (%) | Not relevant (%) | Very Important (%) | Important (%) | Useful (%) | Sometimes useful (%) | Not relevant (%) |
| More severe diarrhoea | 14.8 | 55.6 | 7.4 | 20.4 | 1.9 | 17.2 | 51.7 | 10.3 | 17.2 | 3.4 | 12.0 | 60.0 | 4.0 | 24.0 | 0.0 |
| Prolonged diarrhoea | 53.6 | 42.9 | 3.6 | 0.0 | 0.0 | 51.7 | 41.4 | 6.9 | 0.0 | 0.0 | 55.6 | 44.4 | 0.0 | 0.0 | 0.0 |
| Other symptoms in addition to diarrhoea e.g. vomiting... | 7.1 | 23.2 | 32.1 | 28.6 | 8.9 | 3.4 | 17.2 | 44.8 | 27.6 | 6.9 | 11.1 | 29.6 | 18.5 | 29.6 | 11.1 |
| Pre-existing illness | 1.8 | 40.0 | 23.6 | 30.9 | 3.6 | 0.0 | 32.1 | 28.6 | 39.3 | 0.0 | 3.7 | 48.1 | 18.5 | 22.2 | 7.4 |
| Age extreme (child or elderly) | 14.3 | 48.2 | 16.1 | 19.6 | 1.8 | 13.8 | 41.4 | 13.8 | 27.6 | 3.4 | 14.8 | 55.6 | 18.5 | 11.1 | 0.0 |
| Physical ability rather than disability | 1.8 | 1.8 | 27.3 | 36.4 | 32.7 | 3.4 | 0.0 | 24.1 | 44.8 | 27.6 | 0.0 | 3.8 | 30.8 | 26.9 | 38.5 |
| Mental ability rather than disability | 1.8 | 7.3 | 23.6 | 30.9 | 36.4 | 0.0 | 10.3 | 24.1 | 31.0 | 34.5 | 3.8 | 3.8 | 23.1 | 30.8 | 38.5 |
| Recent travel within U.K. | 0.0 | 7.1 | 12.5 | 46.4 | 33.9 | 0.0 | 13.8 | 10.3 | 37.9 | 37.9 | 0.0 | 0.0 | 14.8 | 55.6 | 29.6 |
| Recent foreign travel | 66.1 | 32.1 | 1.8 | 0.0 | 0.0 | 65.5 | 34.5 | 0.0 | 0.0 | 0.0 | 66.7 | 29.6 | 3.7 | 0.0 | 0.0 |
| Unemployment | 0.0 | 0.0 | 8.9 | 16.1 | 75.0 | 0.0 | 0.0 | 13.8 | 17.2 | 69.0 | 0.0 | 0.0 | 3.7 | 14.8 | 81.5 |
| High status employment | 0.0 | 0.0 | 8.9 | 14.3 | 76.8 | 0.0 | 0.0 | 10.3 | 17.2 | 72.4 | 0.0 | 0.0 | 7.4 | 11.1 | 81.5 |
| Higher skill attainment or education level | 0.0 | 3.6 | 3.6 | 16.1 | 76.8 | 0.0 | 3.4 | 3.4 | 24.1 | 69.0 | 0.0 | 3.7 | 3.7 | 7.4 | 85.2 |
| Greater family or caring commitment | 1.8 | 16.1 | 26.8 | 33.9 | 21.4 | 0.0 | 20.7 | 34.5 | 27.6 | 17.2 | 3.7 | 11.1 | 18.5 | 40.7 | 25.9 |
| Greater social or recreational commitment | 0.0 | 1.9 | 20.4 | 33.3 | 44.4 | 0.0 | 0.0 | 32.1 | 21.4 | 46.4 | 0.0 | 3.8 | 7.7 | 46.2 | 42.3 |
| Patient living in an affluent area | 0.0 | 0.0 | 3.6 | 14.3 | 82.1 | 0.0 | 0.0 | 0.0 | 20.7 | 79.3 | 0.0 | 0.0 | 7.4 | 7.4 | 85.2 |
| Patient living in a poor area | 3.6 | 3.6 | 10.7 | 12.5 | 69.6 | 3.4 | 3.4 | 13.8 | 10.3 | 69.0 | 3.7 | 3.7 | 7.4 | 14.8 | 70.4 |
| Existence of similar cases (e.g. Potential outbreak) | 51.8 | 37.5 | 8.9 | 1.8 | 0.0 | 51.7 | 37.9 | 10.3 | 0.0 | 0.0 | 51.9 | 37.0 | 7.4 | 3.7 | 0.0 |




Heat map showing the importance of risk factors in GPs decisions on requesting stool samples for all GPs and for those serving in practices with catchments predominantly SIMD1 (most deprived, n = 29) or SIMD5 (least deprived, n = 27).
Heat map percentages are colour coded as denoted in the scale ranging from 100% (red) to 0% (green) (Q2 in GP questionnaire, Annex 3.2). In the first column the factors in red text are classified of high importance for a GP to ask for a stool sample (i.e. the sum of GP responses of "very important" & "important" >50%), the factors in blue text are classified of low importance (i.e. the sum of "sometimes useful" & "not relevant" >50%) and the factors in black are classified as inconclusive (both above criteria <50%).

**Figure 3.5 Heat map showing, "in the GPs opinion", the reported importance of factors for stool sample submission by patients.**

| | All GPs | | | | | GPs in most deprived practices | | | | | GPs in least deprived practices | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3. For such a patient that you have asked to submit a stool sample, please rank the 5 factors that you think would most influence them to actually submit a stool sample. | Rank1 (%) | Rank2 (%) | Rank3 (%) | Rank4 (%) | Rank5 (%) | Rank1 (%) | Rank2 (%) | Rank3 (%) | Rank4 (%) | Rank5 (%) | Rank1 (%) | Rank2 (%) | Rank3 (%) | Rank4 (%) | Rank5 (%) |
| More severe diarrhoea | 31.7 | 17.1 | 24.4 | 4.9 | 7.3 | 36.4 | 22.7 | 9.1 | 4.5 | 13.6 | 26.3 | 10.5 | 42.1 | 5.3 | 0.0 |
| Prolonged diarrhoea | 36.6 | 22.0 | 17.1 | 9.8 | 4.9 | 36.4 | 18.2 | 22.7 | 9.1 | 0.0 | 36.8 | 26.3 | 10.5 | 10.5 | 10.5 |
| Other symptoms in addition to diarrhoea e.g. vomiting… | 2.4 | 7.3 | 4.9 | 19.5 | 4.9 | 4.5 | 9.1 | 0.0 | 9.1 | 4.5 | 0.0 | 5.3 | 10.5 | 31.6 | 5.3 |
| Pre-existing illness | 4.9 | 0.0 | 4.9 | 2.4 | 14.6 | 0.0 | 0.0 | 9.1 | 0.0 | 4.5 | 10.5 | 0.0 | 0.0 | 5.3 | 26.3 |
| Age extreme (child or elderly) | 0.0 | 9.8 | 4.9 | 12.2 | 12.2 | 0.0 | 13.6 | 4.5 | 4.5 | 18.2 | 0.0 | 5.3 | 5.3 | 21.1 | 5.3 |
| Physical ability rather than disability | 0.0 | 0.0 | 0.0 | 0.0 | 2.4 | 0.0 | 0.0 | 0.0 | 0.0 | 4.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Mental ability rather than disability | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Recent travel within U.K. | 0.0 | 0.0 | 2.4 | 0.0 | 4.9 | 0.0 | 0.0 | 4.5 | 0.0 | 9.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Recent foreign travel | 22.0 | 26.8 | 26.8 | 9.8 | 4.9 | 18.2 | 22.7 | 27.3 | 18.2 | 4.5 | 26.3 | 31.6 | 26.3 | 0.0 | 5.3 |
| Unemployment | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| High status employment | 2.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 5.3 | 0.0 | 0.0 | 0.0 | 0.0 |
| Higher skill attainment or education level | 0.0 | 4.9 | 0.0 | 4.9 | 2.4 | 0.0 | 0.0 | 0.0 | 0.0 | 4.5 | 0.0 | 10.5 | 0.0 | 10.5 | 0.0 |
| Greater family or caring commitment | 0.0 | 0.0 | 4.9 | 7.3 | 7.3 | 0.0 | 0.0 | 4.5 | 13.6 | 9.1 | 0.0 | 0.0 | 5.3 | 0.0 | 5.3 |
| Greater social or recreational commitment | 0.0 | 0.0 | 0.0 | 4.9 | 0.0 | 0.0 | 0.0 | 0.0 | 4.5 | 0.0 | 0.0 | 0.0 | 0.0 | 5.3 | 0.0 |
| Patient living in an affluent area | 0.0 | 0.0 | 0.0 | 12.2 | 7.3 | 0.0 | 0.0 | 0.0 | 13.6 | 9.1 | 0.0 | 0.0 | 0.0 | 10.5 | 5.3 |
| Patient living in a poor area | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Existence of similar cases (e.g. Potential outbreak) | 0.0 | 7.3 | 7.3 | 12.2 | 24.4 | 0.0 | 9.1 | 13.6 | 22.7 | 18.2 | 0.0 | 5.3 | 0.0 | 0.0 | 31.6 |




Heat map showing, "in the GPs opinion" (Q3 in GP questionnaire, Annex 3.2), the reported importance of factors for stool sample submission by patients. This is provided for all GPs and for those serving in practices with catchments predominantly SIMD1 (most deprived, n = 29) or SIMD5 (least deprived, n = 27). Heat map percentages are colour coded as denoted in the scale ranging from 50% (red) to 0% (green). In the first column the factors in red text are classified of high importance to submit a stool sample (i.e. the sum of the ranks >80%) and the factors in blue text are classified of low importance (i.e. the sum of the ranks <80%).

Question 3 asks the GP for their opinion on the reported importance (ranked 1 to 5) of factors for stool sample submission by patients. There were 56 GPs who answered the question (Figure 3.5), however only 41 answered according to what was asked (i.e. ranking 1 to 5). This was because there was some confusion in how the question should be answered. Hence only those 41 who answered it as required were included in the analysis.

Thirty-seven percent of the GPs considered "prolonged diarrhoea" as the most important factor to a patient when deciding to submit a stool sample. This is followed by "more severe diarrhoea" (32%) and "recent foreign travel" (22%) (Figure 3.5). Socio-economic factors (e.g. "patient living in a poor area" or "unemployment") are considered not relevant (100%) for patients when they decide to submit a stool sample. The heat map patterns indicate that GPs' estimates of impact of factors prompting sample submission are independent of patient's level of deprivation (Figure 3.5). This is consistent with the Mann-Whitney's non parametric test ($P>0.05$ when comparing risk factors between both deprivation categories).

## 3.3.2 Practice Team Information (PTI) study

### 3.3.2.1 Aims

The aim of this study was to determine the number and incidence of GI diagnoses at the GP level, stratified by deprivation and considering the effects of age and gender.

### 3.3.2.2 PTI data and methods

This Practice Team Information (PTI) study (http://www.isdscotland.org/Health-Topics/General-Practice/GP-Consultations/What-is-PTI.asp) has recorded the diagnoses of patients presenting at 58 GP practices across Scotland from September 2011 to August 2013. It also recorded the proportion of the population in each practice by deprivation quintile. These data are stored by the Electronic Data Research and Innovation Service (eDRIS) at the Information Services Division (ISD) of NHS Scotland. In total >10 million records are stored that detail the illness/diagnoses of the patients.

At the point when the patient presents, the GP is unlikely to know that the patient has campylobacteriosis rather than an infection with any other infectious agent. GP behaviour at the stage of presentation with gastroenteritis is key here – diagnosis of campylobacteriosis requires laboratory stool testing. Therefore the current study focussed on the RCG3 diagnosis "Gastroenteritis of possible infectious origin" and associated patient metadata (age, gender, data zone, date of diagnosis, GP's practice name).

An email request was sent by eDRIS to the 58 PTI practices asking permission to utilise the above data for the current study. An email reminder was sent to those who did not respond, followed by a phone call. The University of Aberdeen was not allowed to approach the practices directly. In total 43 (74%) practices

provided consent and these served 258,292 individuals (4.9% of the Scottish population) (Figure 3.6). It is stated that these "are broadly representative of the Scottish population in terms of age, gender, deprivation and urban/rural mix" (https://www.isdscotland.org/Health-Topics/General-Practice/Publications/2013-10-29/2013-10-29-PTI-Report.pdf). The PTI data were linked with data obtained from MMDLs and this is explained in section 3.3.4.

Table 3.3 provides a summary of the acquired PTI data. It contains the number of participating GP practices in the PTI study by health board, together with the number of practices giving consent for their data to be used in the present study. The last column gives the number of RCG3 diagnoses (n = 1092) originating from those practices that consented.

**Figure 3.6 The geographical distribution of the PTI GP practices participating in this study.**

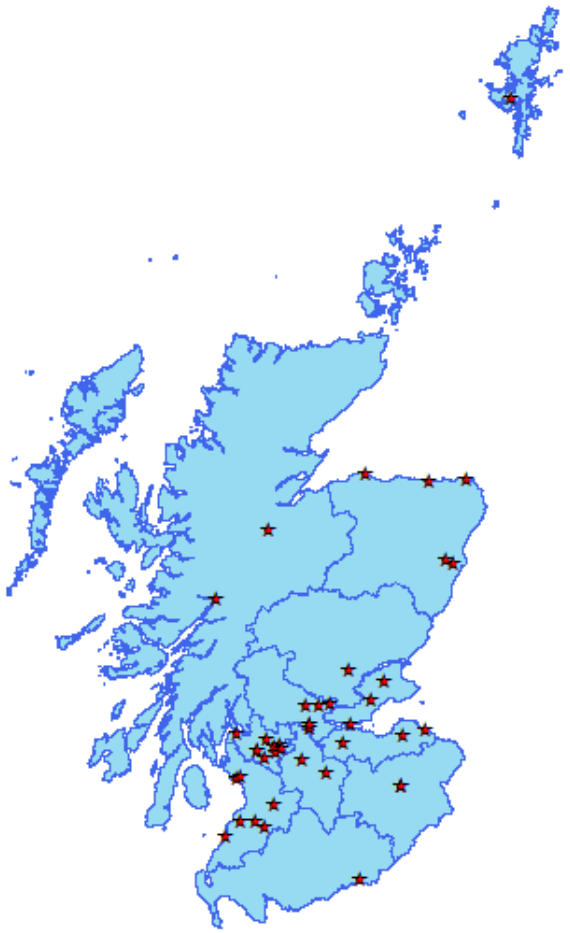


**Table 3.3 Summary of the PTI data obtained from the GP practices.**

| Health board | Number of GP practices in PTI study | Number of GP practices in PTI consenting to participate in current study | Number of RCG3 diagnoses from participating PTI GP practices |
|---|---|---|---|
| Ayrshire & Arran | 7 | 7 | 124 |
| Borders | 3 | 3 | 12 |
| Fife | 5 | 3 | 151 |
| Forth Valley | 7 | 6 | 229 |
| Grampian | 6 | 5 | 97 |
| Greater Glasgow & Clyde | 12 | 9 | 206 |
| Highland | 4 | 2 | 6 |
| Lanarkshire | 3 | 2 | 75 |
| Lothian | 5 | 3 | 142 |
| Tayside | 2 | 1 | 3 |
| Dumfries & Galloway | 2 | 1 | 42 |
| Shetland | 2 | 1 | 5 |
| Grand Total | 58 | 43 | 1092 |

The percentage of RCG3 diagnoses and bootstrapped 95% confidence intervals were calculated for each deprivation SIMD quintile (Manly 2007). Two different methods were used to establish a proxy for the socio-economic status of the patients: (i) the socio-economic status of the patients was given by the SIMD quintile of their data zone of residence; (ii) the population distribution by SIMD quintiles in each GP practice was used to probabilistically attribute an SIMD quintile to each patient. In both cases Monte Carlo simulations (n=10,000) were used to calculate the average percentage of RCG3 diagnoses and 95% confidence intervals for each SIMD quintile. This was performed in Excel using the @Risk7.0.1 add-in (http://www.palisade.com/).

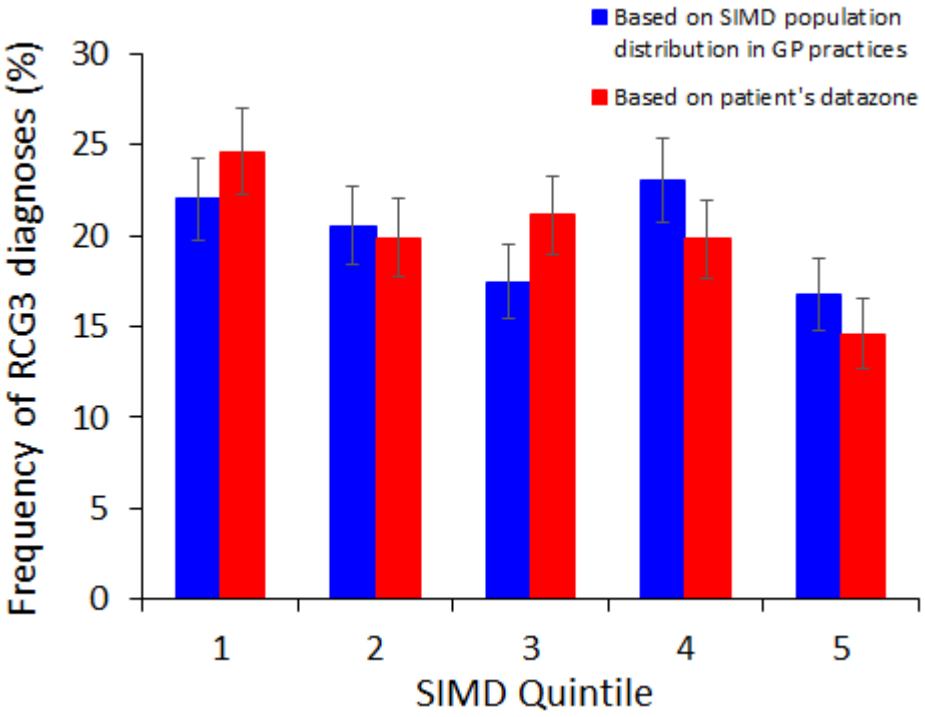
The average incidence (diagnoses /100,000 population) of RCG3 diagnoses and 95% bootstrap confidence intervals were then calculated for each deprivation SIMD quintile. Finally the distribution of RCG3 diagnoses by age and gender were determined and compared with the national population distribution.

### 3.3.2.3 Results and Discussion from PTI study

Figure 3.7 presents the percentage of RCG3 diagnoses by SIMD quintile. The results show that there were significantly (P<0.05) fewer diagnoses in the SIMD5 quintile (16.8%), compared with SIMD1 (22.1%) based on SIMD population distribution within the practices. Similar results (14.6% in SIMD5 vs. 24.6% in SIMD1) were obtained when the patient's data zone was used to establish the SIMD quintile.

This trend runs counter to the *Campylobacter-specific* case gradient by deprivation and is more in keeping with the commonly found socio-economic gradient of disease incidence (deprivation generally correlates with disease impact) (www.scotpho.org.uk/media/1656/sbod2016-deprivation-report-aug18.pdf).

**Figure 3.7 Frequency of RCG3 diagnoses from the PTI study stratified by deprivation quintile**

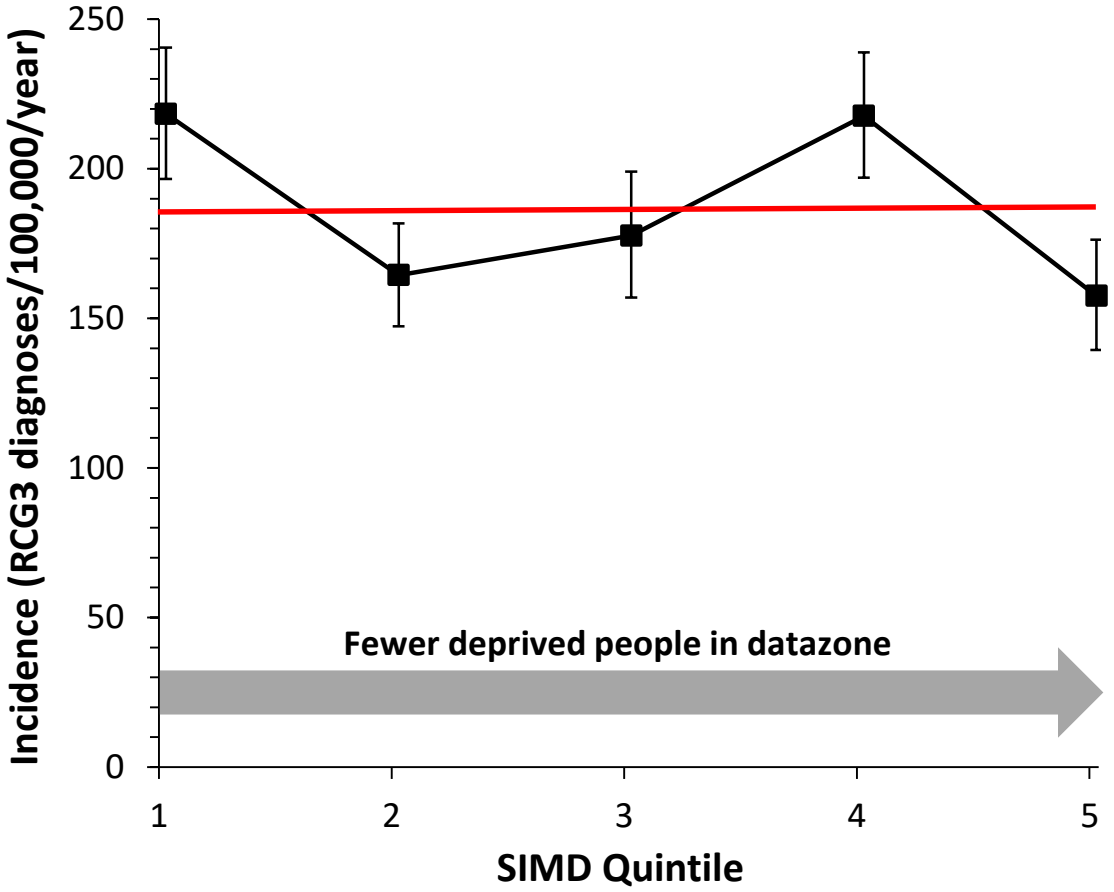



Frequency of RCG3 diagnoses ('Gastroenteritis of possible infectious origin') from the PTI study stratified by deprivation quintile (error bars represent 95% bootstrapped confidence intervals).

The average incidence rate for RCG3 diagnoses during Aug 2011-Sep 2013 was 186.2 diagnoses/100,000/year for the population in the GP practices in the study. Figure 3.8 presents the incidence of RGC3 diagnoses by SIMD quintile and shows that the incidence was significantly (P<0.05) lower among people in the least deprived quintile, than the most deprived (157.5 compared with 218.3 diagnoses/100,000). There are a number of possible reasons for this which include: a lower proportion of people from SIMD5 presenting to a GP when they

have a GI infection; exposure to GI pathogens may be different for the SIMD1 and SIMD5 populations and people in the SIMD5 deprivation quintile may be less susceptible to gastrointestinal infections. This is something that was unknown previously. It helps understand where issues in reporting occur. If this was known prior to the study then the approach would have been modified to account for it.

**Figure 3.8 Incidence of RCG3 diagnoses from the PTI study by deprivation quintile**



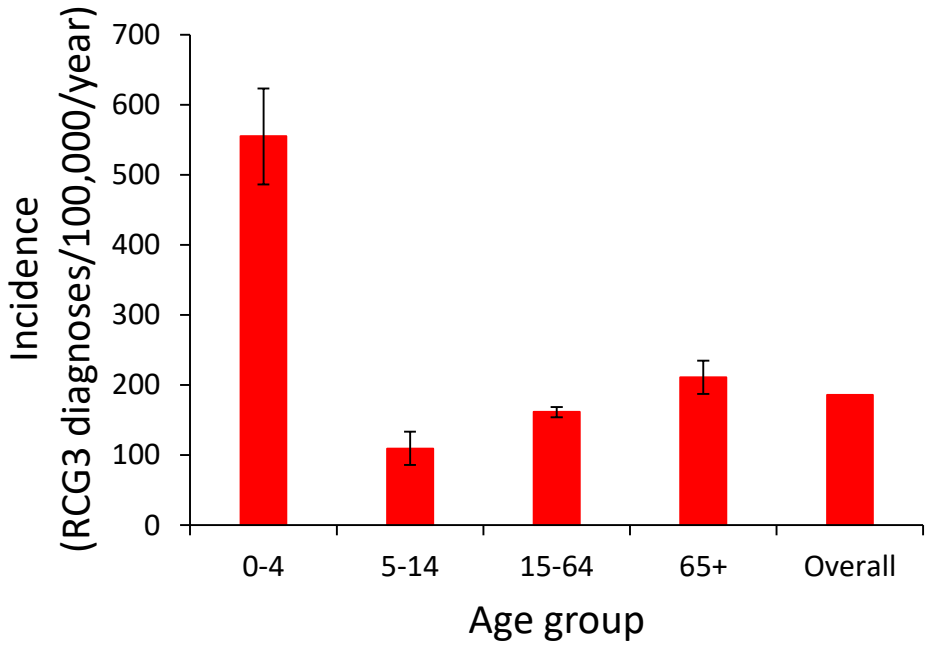

Incidence of RCG3 diagnoses ('Gastroenteritis of possible infectious origin') from the PTI study by deprivation quintile (error bars represent 95% bootstrapped confidence intervals). The red horizontal line represents the average incidence.

Figure 3.9(a) shows that the incidence of RCG3 diagnoses is significantly ($P<0.05$) higher in children <5 years old compared with all other age groups (it is three times higher than the overall average incidence). Also, the distribution of RCG3 diagnoses by age shows that there was a significantly ($P<0.05$) higher percentage (16.5%) of children (0-4 years) diagnosed than the corresponding proportion of this age group in the Scottish population (5.5%) (Figure 3.9(b)). This is not necessarily surprising as one might expect parents to present their children to the GP with a GI illness more often than the adult population. Lower

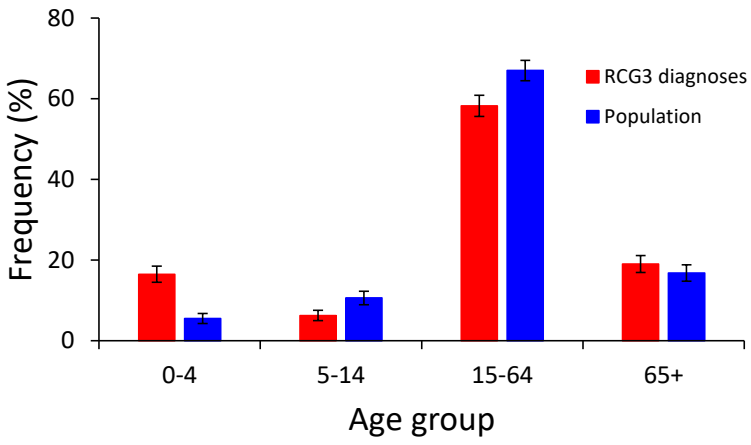
immunity in children may also play a role. For example children <5 years old are more susceptible to norovirus than older children and adults (Simmons, Gambhir et al. 2013). The opposite is the case for the 5-15 and 15-64 years old, where the percentages of RCG3 diagnoses are significantly (P<0.05) lower (6.2 % and 58.2%) than the corresponding percentages of these population groups in the Scottish population (10.6% and 67.1%). There is no difference in the elderly group (19% GI diagnosed with 17% of the Scottish population being 65+ years old).

**Figure 3.9 (a) The incidence of RCG3 diagnoses by age and (b) The distribution of RCG3 diagnoses by age compared with the age distribution of the Scottish population.**

(a)

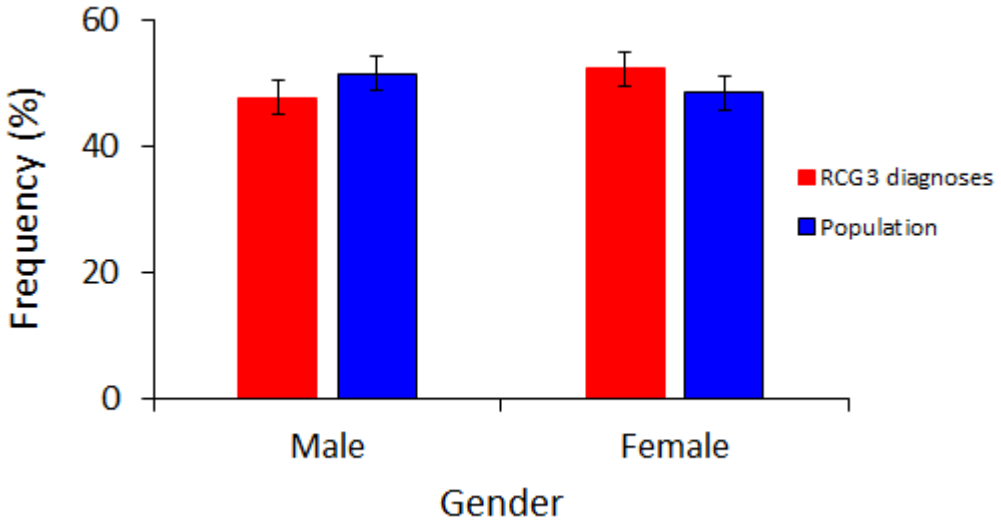



(b)



There is no significant (P>0.05) difference in the distribution of RCG3 diagnoses by gender (47.7±2.7% male, 52.3±2.7% female) (Figure 3.10). This follows the male/female distribution of the Scottish population. In terms of incidence the rates in males and females are also the same (183.2±10.4 diagnoses/100,000/year in male, 189.0±9.8 diagnoses/100,000/year in female).

**Figure 3.10 The distribution of RCG3 diagnoses by gender compared with the gender distribution of Scottish population.**




## 3.3.3 Medical microbiological diagnostic laboratories (MMDLs) study

### 3.3.3.1 Aims

The aim of this study was to determine the proportion of patients diagnosed with a GI infection by their GP that end up with a stool sample being received for analysis at the medical microbiology diagnostic laboratories (MMDLs).

### 3.3.3.2 MMDL data and methods

Non-patient identifiable data (age, gender, data zone, date sample received, GP's practice name, *Campylobacter*/other GI pathogen result) from stool samples received by the MMDL labs over the two years (Sep 2011 – Aug 2013) that coincided with the PTI study were collated. The island health boards were not included because there was no IT service to extract the data from Shetland and Orkney and the Western Isles have small populations and no PTI practices which would be required for data linkage (see below).

Table 3.4 presents the summary of the data return by health board. For ten health boards (AA, BR, FF, FV, HI, GR, GC, LA, LO and TY) data were complete. Data were incomplete (e.g. missing postcodes and/or missing data and/or missing data for particular time periods) for DG health board and it was therefore not possible to perform data linkage with the PTI study.

The MMDL and PTI data were linked based on the following five descriptors: date (date recorded in PTI and date received in MMDL); GP practice name; data zone of patient, age and sex.

The percentage of linkage between the PTI and MMDL datasets was used to determine the under reporting factor between the following steps in the reporting pyramid - "GP diagnosis of GI infection" to "Stool sample received by MMDL" (Figure 3.1). This was calculated for each health board and each deprivation quintile. Statistical significance was determined by Fisher's exact test (Fisher 1935).

#### 3.3.3.3 Results and Discussion from MMDL study

Table 3.4 provides the summary statistics of the data linkage between MMDL and PTI studies. Only a fraction (11.3%) of RCG3 diagnoses for the 10 health boards where data are complete have samples linked to the MMDL. There are a number of potential reasons for this: (1) only some GPs from a practice may have participated in the PTI study; (2) some GPs may have not recorded the RCG3 diagnoses; (3) some MMDL samples may be from individuals at hospital; (4) there may be duplicate samples in the MMDL data (there is no ID number of the patient in the MMDL data to perform complete de-duplication); (5) patients with an RCG3 diagnosis may not have been asked for a stool sample by their GP and (6) patients may not have provided a stool sample when requested by their GP.

Table 3.4 also shows that the percentage of RCG3 diagnoses reported in the PTI study which are found in the MMDLs varies by health board. For example the extremes (BO (0%) and AA (23.4%) are significantly different to each other ($P<0.05$ by Fisher's exact test). TY and HG were not compared due to the small amount of data available.

There were 22 and 17 MMDL stool samples from the SIMD1 and SIMD5 quintiles corresponding to incidences of 23.6 and 17.4 /100,000 respectively. However this was not significantly different as the bootstrapped confidence intervals overlapped.

Comparison of the PTI/MMDL reporting ratios between each of the SIMD quintiles (Table 3.5) was performed using Fisher's exact test. However, no significant differences were observed between deprivation quintiles ($P>0.05$).

**Table 3.4 Linkage between MMDL and PTI data by Health Board**

| Health boards with complete data | Number of GP practices in PTI study | Number of GP practices in PTI consenting to participate in current study | Number of RCG3 diagnoses from participating PTI GP practices | Total No stool samples recorded by MMDL originating from GP practices participating in PTI study | Number of linked faecal samples tested in MMDL also with RCG3 diagnosis in PTI | MMDL/PTI (%) |
|---|---|---|---|---|---|---|
| Ayrshire & Arran | 7 | 7 | 124 | 1047 | 29 | 23.4% |
| Borders | 3 | 3 | 12 | 524 | 0 | 0% |
| Fife | 5 | 3 | 151 | 982 | 18 | 11.9% |
| Forth Valley | 7 | 6 | 229 | 1502 | 11 | 4.8% |
| Grampian | 6 | 5 | 97 | 1494 | 19 | 19.6% |
| Greater Glasgow & Clyde | 12 | 9 | 206 | 386 | 12 | 5.8% |
| Highland | 4 | 2 | 6 | 143 | 0 | 0% |
| Lanarkshire | 3 | 2 | 75 | 214 | 16 | 21.3% |
| Lothian | 5 | 3 | 142 | 728 | 13 | 9.2% |
| Tayside | 2 | 1 | 3 | 250 | 0 | 0% |
| Total | 54 | 41 | 1045 | 7270 | 118 | 11.3% |

| | | | | | | |
|---|---|---|---|---|---|---|
| Health board with incomplete data | | | | | | |
| Dumfries & Galloway | 2 | 1 | 42 | 273 | data incomplete | data incomplete |
| Shetland | 2 | 1 | 5 | not available | not available | not available |
| Total | 4 | 2 | 47 | 273 | - | - |
| Grand Total | 58 | 43 | 1092 | 7543 | 118 | NA |

Sep 2011- Aug 2013

**Table 3.5 Linkage between MMDL and PTI data for the ten health boards**

| SIMD quintile | Number of RCG3 PTI diagnoses | Population in contributing GP practices | RCG3 PTI diagnoses in the population per year (%) | Number of corresponding stool samples in MMDL | Under-reporting factor between PTI and MMDL |
|---|---|---|---|---|---|
| 1 (Fewer deprived people in ↓) | 246 | 46551 | 0.26 | 22 | 11.2 |
| 2 | 194 | 48903 | 0.20 | 31 | 6.3 |
| 3 | 190 | 42274 | 0.23 | 22 | 8.6 |
| 4 | 248 | 47113 | 0.26 | 26 | 9.5 |
| 5 | 166 | 48957 | 0.17 | 17 | 9.8 |
| Total | 1044 | 233798 | 0.22 | 118 | 8.9 |

Only where there is complete data. Stratified by SIMD deprivation quintile. MMDL and PTI comprise two years of data (September 2011 – August 2013).

## 3.4 Estimation of reporting biases at the reported case level

Biases in human campylobacteriosis reporting may be observed in data at the reported case level. For example if there was a greater tendency to report foreign travel cases then this may explain differences between reporting rates from the most and least deprived deprivation quintiles. Further, it is also possible that there may be variation by deprivation in likelihood to attend their GP when symptoms are mild because of variable requirement to provide a "Fitness for work" note depending on employment status. It is also possible that people who do not work (e.g. because have the financial means to support themselves or unemployed) may have more time available to attend the doctor. This section looks for biases at the reported case level and within hospitalised cases which would be expected to include the most severe symptoms.

### 3.4.1 Aims

To determine if the spectrum of symptoms in reported cases and hospitalisations varies by deprivation.

To determine whether the spectrum of symptoms varies between hospitalised and not-hospitalised cases.

### 3.4.2 Reported case level data and methods

Estimation of the reporting biases at the reported case level is based on the case questionnaires from the case-control study (questionnaire is at Annex 2.1). The

case patients (n=598 of which 590 had SIMD information) were asked the following questions in section 3 of the questionnaire:

- when did they first started to feel unwell,
- when did they see the doctor,
- the start and duration of symptoms,
- if they were admitted to hospital (and for how long) and
- how many other people with similar symptoms were in the household.

Frequency distributions of each symptom reported was generated for SIMD1 and SIMD5. Then, frequency distributions were produced of the duration of symptoms stratified by deprivation (i.e. SIMD1 and SIMD5).

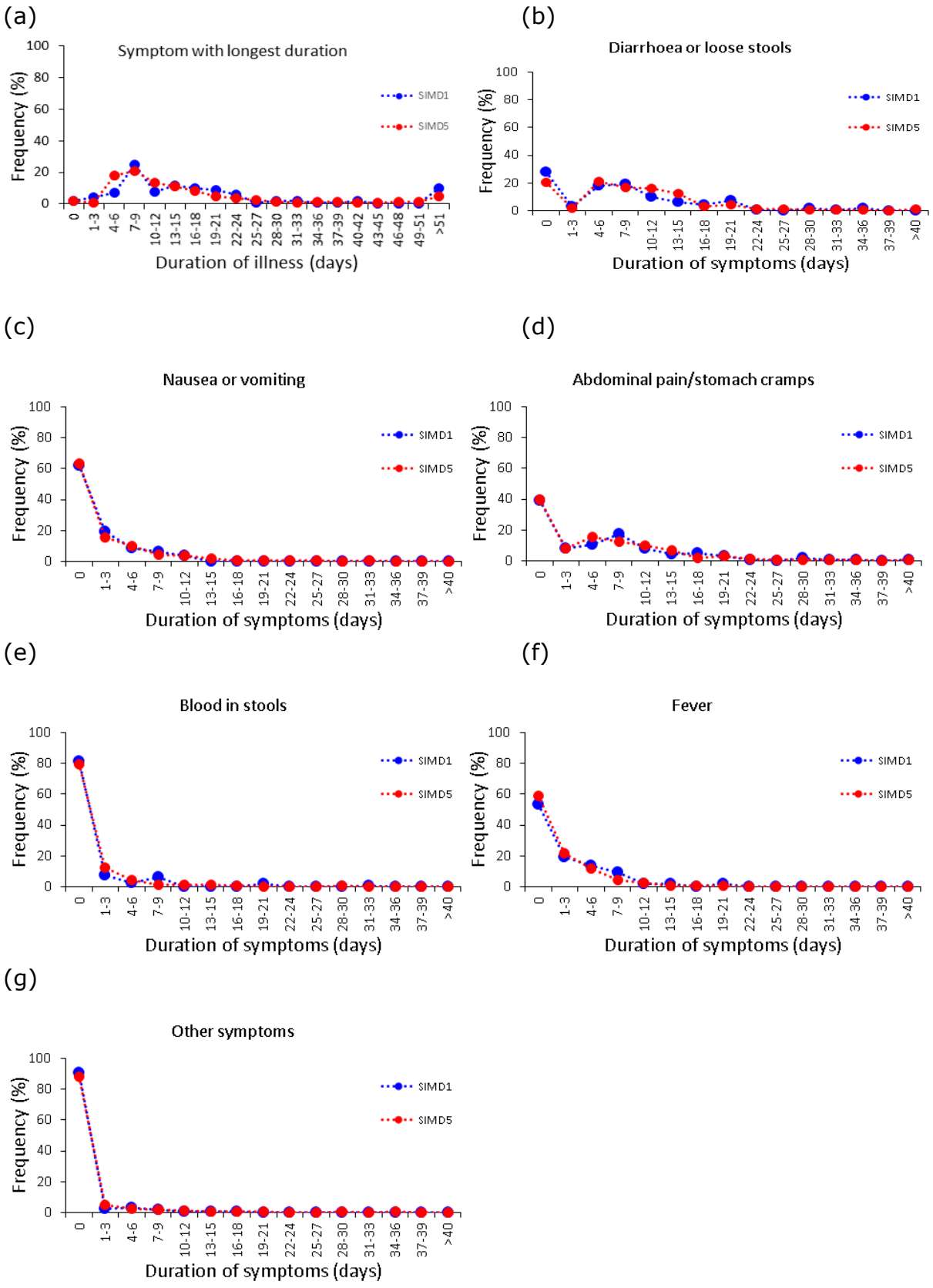
For the cases that were hospitalised the frequency of symptoms by deprivation (SIMD quintiles 1 and 5) was calculated.

Univariate and multivariate logistic regression (Cox 1958) was used to quantify differences between hospitalised and not-hospitalised cases by the following risk factors: symptom; SIMD; gender age; and travel out with Scotland.

### 3.4.3 Reported case level results and discussion

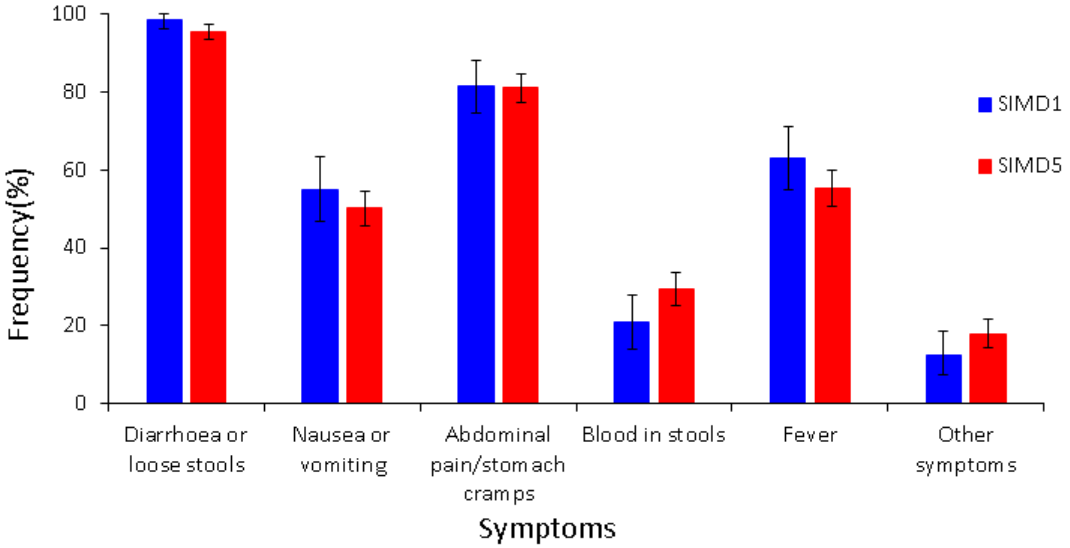
The duration of each of the case symptoms (n=590) did not vary by deprivation (Figure 3.11). Further, the frequency of each symptom did not vary by deprivation (Figure 3.12). It can also be observed in Figure 3.11 that 6% of cases had symptoms lasting >51 days. It is possible that some of these cases had other health problems which may have been the reason for the long duration. Blood in stools was not uncommon (23.1% of cases had this symptom) but only 3.1% suffered from this for >9 days. The "other symptoms" provided were predominantly: headache; weakness; tiredness; muscle and joint pain.

**Figure 3.11 Duration of illness by deprivation for the cases from the case-control study**



Duration of illness by deprivation (SIMD1 and SIMD5) for the cases from the case-control study: (a) symptom with longest duration; (b) diarrhoea or loose stools; (c) nausea or vomiting; (d) abdominal pain/stomach cramps; (e) blood in stools; (f) fever and (g) other symptoms. Note there were 135 reported cases from SIMD1 and 455 from SIMD5. Zero days includes those cases that did not report duration.

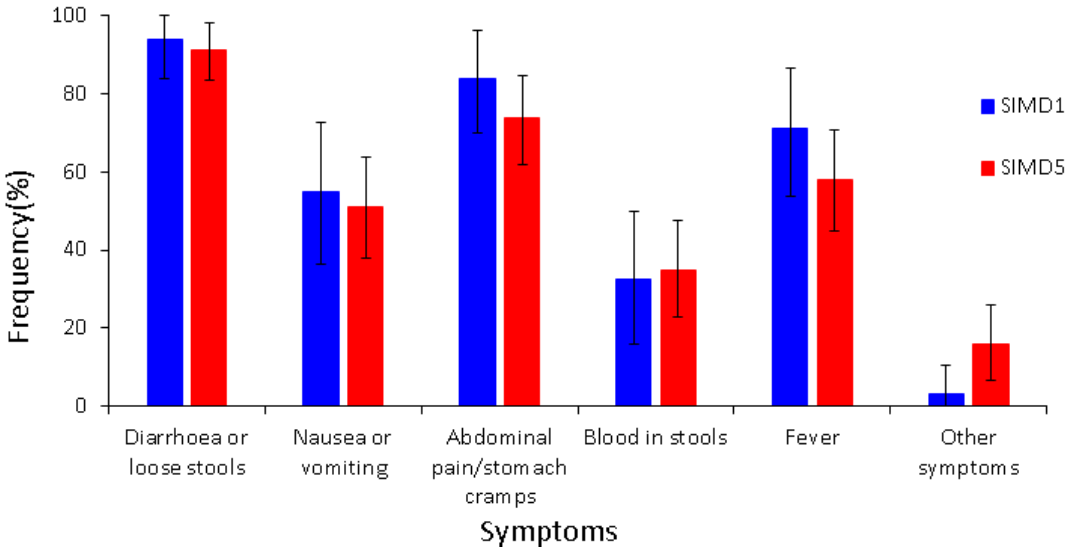
**Figure 3.12 Frequency of symptoms by deprivation for the cases from the case-control study**




Frequency of symptoms by deprivation (SIMD1 and SIMD5 quintiles) for the cases from the case-control study.

There were 88 cases from the case-control study that were hospitalised (57 from SIMD5 and 31 from SIMD1). There was no difference in the frequency of symptoms stratified by deprivation (Figure 3.13). The "other symptoms" provided by the hospitalised cases were predominantly the same as the not-hospitalised cases (e.g. headache, tiredness and weakness). There were a couple of other symptoms (or possibly consequences) provided that were very fast heart rate and falling.

**Figure 3.13 Frequency of symptoms by deprivation for the hospitalised cases from the case-control study**




Frequency of symptoms by deprivation (SIMD quintiles 1 and 5) for the hospitalised cases from the case-control study.

Comparing hospitalised with not-hospitalised cases from the case-control study showed in the univariate analysis that proportionally higher numbers were hospitalised in SIMD1 (most deprived) compared with SIMD5 (least deprived) (Table 3.6). It was also found that proportionally fewer cases that were hospitalised (81/88 =92%) had diarrhoea or loose stools than not-hospitalised (494/502 = 98%). It is surprising that there were some hospitalised and not-hospitalised cases that did not have diarrhoea or loose stools because it would be expected that this symptom would be a pre-requisite for a stool sample to be taken. However, the range of symptoms given by these individuals included stomach cramps and nausea or vomiting so it is possible a stool sample was taken because of these other symptoms or that the symptoms were not recorded accurately.

**Table 3.6 Univariate logistic regression comparing hospitalised and not-hospitalised cases from the case-control study by risk factor**

| Risk Factor | Number of hospitalised cases | Number of not hospitalised cases | OR* | 95%CI** | P-value |
|---|---|---|---|---|---|
| Disease symptoms | | | | | |
| Diarrhoea or loose stools | | | | | |
| No(reference) | 7 | 8 | 1 | | |
| Yes | 81 | 494 | 0.187 | 0.066,0.531 | 0.002 |
| Nausea or vomiting | | | | | |
| No | 42 | 246 | 1 | | |
| Yes | 46 | 256 | 1.052 | 0.669,1.656 | 0.825 |
| Abdominal pain / stomach cramps | | | | | |
| No(reference) | 20 | 84 | 1 | | |
| Yes | 68 | 418 | 0.683 | 0.394,1.185 | 0.175 |
| Blood in stools | | | | | |
| No(reference) | 58 | 366 | 1 | | |
| Yes | 30 | 136 | 1.392 | 0.859,2.256 | 0.179 |
| Fever | | | | | |
| No(reference) | 33 | 218 | 1 | | |
| Yes | 55 | 284 | 1.279 | 0.803,2.039 | 0.300 |

| (continued)<br>Risk Factor | Number of hospitalised cases | Number of not hospitalised cases | OR* | 95%CI** | P-value |
|---|---|---|---|---|---|
| Other symptoms | | | | | |
| No(reference) | 78 | 412 | 1 | | |
| Yes | 10 | 90 | 0.587 | 0.292,1.178 | 0.134 |
| SIMD quintile | | | | | |
| SIMD5(least deprived) (reference) | 57 | 392 | 1 | | |
| SIMD1(most deprived) | 31 | 102 | 2.090 | 1.282,3.407 | 0.003 |
| Gender | | | | | |
| Female(reference) | 38 | 263 | 1 | | |
| Male | 50 | 239 | 1.448 | 0.917,2.286 | 0.112 |
| Age group | | | | | |
| 65+ years old(reference) | 30 | 124 | 1 | | |
| 5-14 years old | <5 | 11 | 0.376 | 0.047,3.025 | 0.358 |
| 15-64 years old | 55 | 348 | 0.653 | 0.400,1.066 | 0.088 |
| Travel outside Scotland | | | | | |
| No(reference) | 53 | 314 | 1 | | |
| Yes | 24 | 158 | 0.900 | 0.536,1.512 | 0.690 |

*OR, odds ratio. **95%CI, 95% confidence interval
Note: For brevity the intercepts are not displayed. Significant results are coloured in red (comparison is significantly higher than the reference) and blue (comparison is significantly lower).

Multivariate analysis was then performed using only those risk factors with P<0.25 in the univariate analysis. It was found that proportionally lower numbers of individuals were hospitalised with diarrhoea or loose stools as was found in the univariate analysis (Table 3.7). Also, there are proportionally higher numbers of individuals hospitalised in SIMD1 (most deprived) compared with SIMD5 (least deprived). This agrees with the univariate analysis and the findings in chapter 5 which shows higher hospitalisation rates in the most deprived areas of Scotland.

**Table 3.7 Multivariate logistic regression comparing hospitalised and not-hospitalised cases from the case-control study by risk factor.**

| Risk Factor | Number of hospitalised cases | Number of not hospitalised cases | OR* | 95%CI** | P-value |
|---|---|---|---|---|---|
| Disease symptoms | | | | | |
| Diarrhoea or loose stools | | | | | |
| No(reference) | 7 | 8 | 1 | | |
| Yes | 81 | 494 | 0.212 | 0.066,0.0681 | 0.009 |
| Abdominal pain / stomach cramps | | | | | |
| No(reference) | 20 | 84 | 1 | | |
| Yes | 68 | 418 | 0.835 | 0.439,1.588 | 0.583 |
| Blood in stools | | | | | |
| No(reference) | 58 | 366 | 1 | | |
| Yes | 30 | 136 | 1.626 | 0.964,2.742 | 0.069 |
| Other symptoms | | | | | |
| No(reference) | 78 | 412 | 1 | | |
| Yes | 10 | 90 | 0.723 | 0.353,1.480 | 0.375 |
| SIMD quintile | | | | | |
| SIMD5(least deprived)(reference) | 57 | 392 | 1 | | |
| SIMD1(most deprived) | 31 | 102 | 2.309 | 1.382,3.857 | 0.001 |

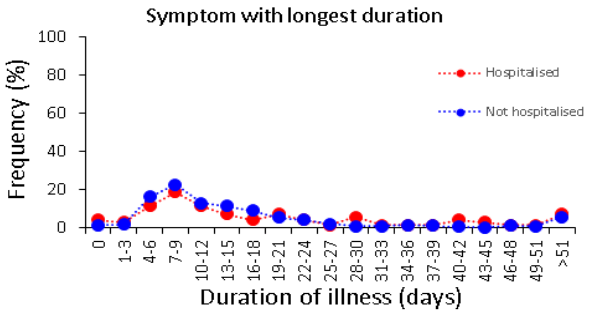
(continued)

| Risk Factor | Number of hospitalised cases | Number of not hospitalised cases | OR* | 95%CI** | P-value |
|---|---|---|---|---|---|
| Gender | | | | | |
| Female(reference) | 38 | 263 | 1 | | |
| Male | 50 | 239 | 0.668 | 0.409,1.091 | 0.107 |
| Age group | | | | | |
| 65+ years old(reference) | 30 | 124 | 1 | | |
| 5-14 years old | <5 | 11 | 0.412 | 0.049,3.445 | 0.413 |
| 15-64 years old | 55 | 348 | 0.606 | 0.351,1.045 | 0.072 |
| Intercept | na*** | na | 1.149 | na, na | 0.821 |

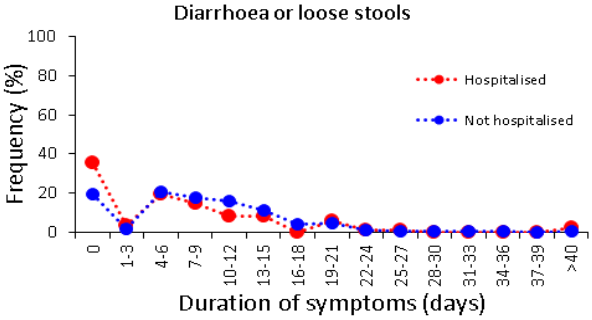
*OR, odds ratio. **95%CI, 95% confidence interval. *** na, not applicable
Significant results are coloured in red (comparison is significantly higher than the reference) and blue (comparison is significantly lower).

## Figure 3.14 Duration of illness by hospitalisation for the cases from the case-control study

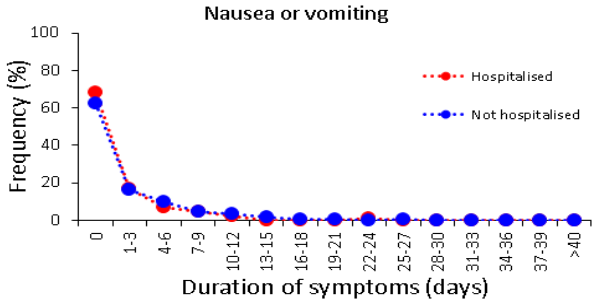
(a)



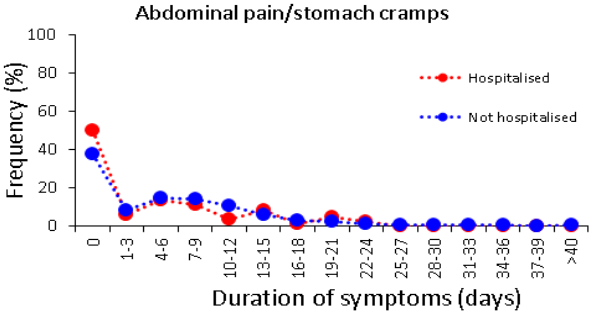

(b)



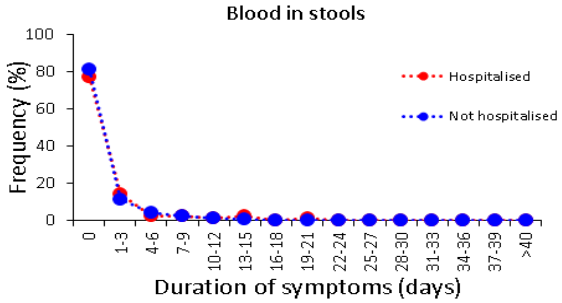

(c)



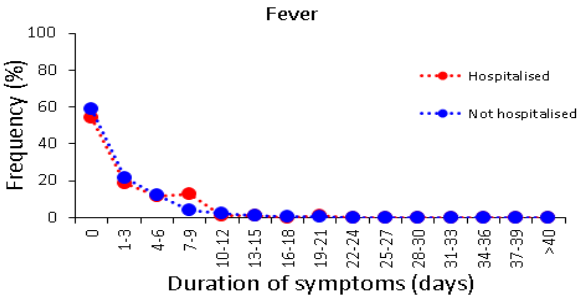

(d)



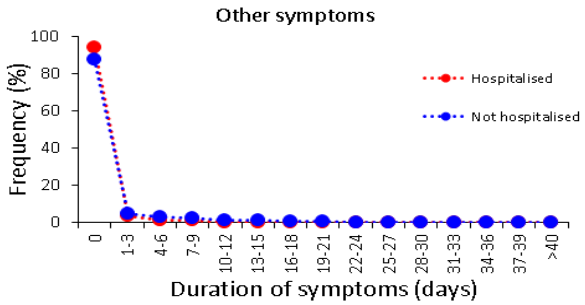

(e)




(f)



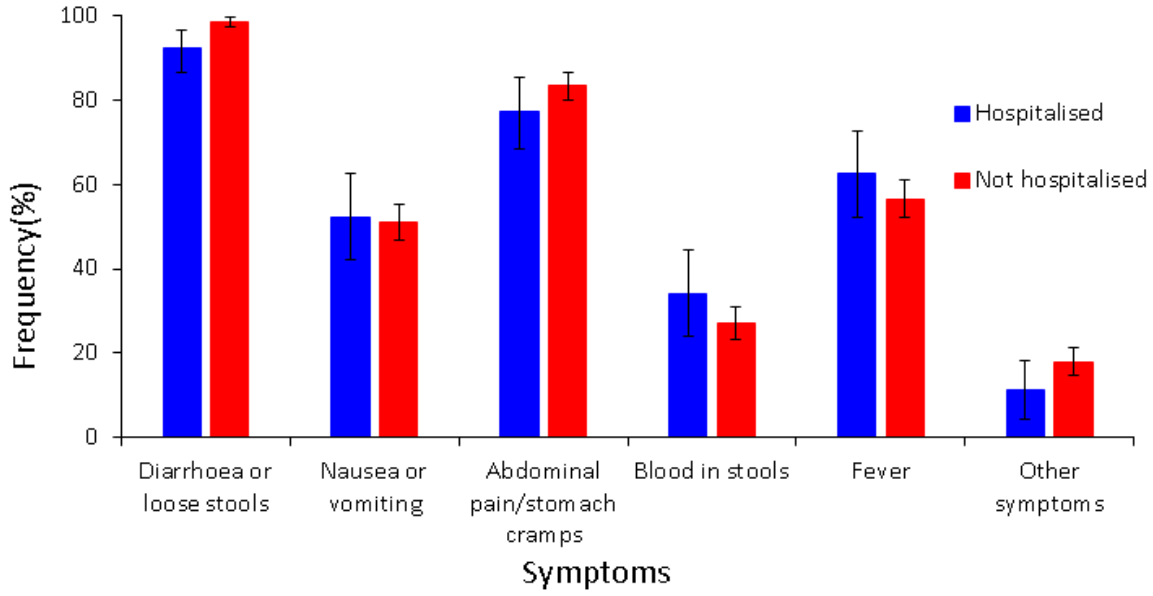

(g)




(a) symptom with longest duration; (b) diarrhoea or loose stools; (c) nausea or vomiting; (d) abdominal pain/stomach cramps; (e) blood in stools; (f) fever and (g) other symptoms. (Note: there were 88 hospitalised and 502 not hospitalised cases. Zero days includes those cases that did not report duration).

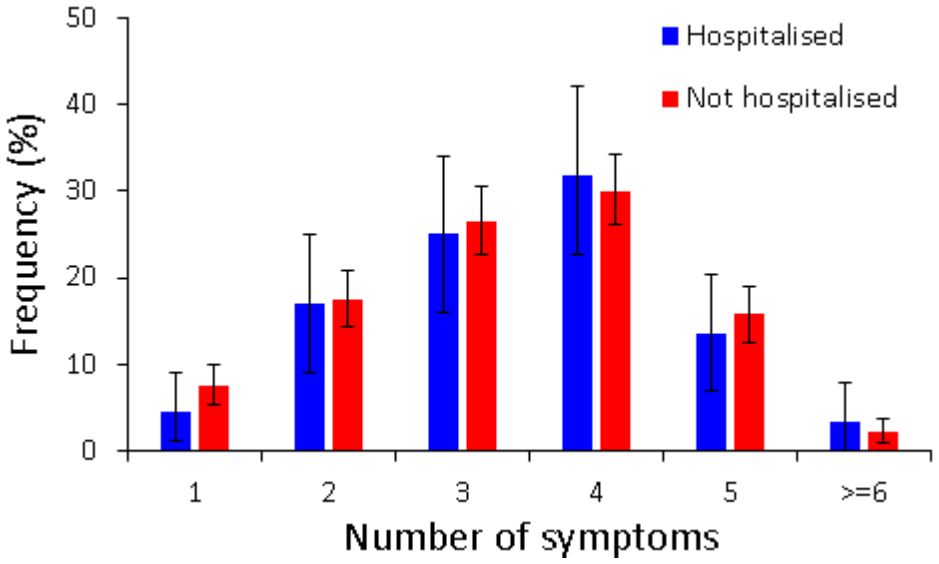
Generally there appears to be little difference between hospitalised/ not hospitalised cases for all symptoms except diarrhoea or loose stools (Figure 3.14 and Figure 3.15) as discussed previously (Table 3.6).

**Figure 3.15 Frequency of symptoms for hospitalised and not hospitalised cases from the case-control study.**




There is no difference in the number of symptoms between hospitalised and not hospitalised cases (Figure 3.16). This provides further evidence from the data collected that it is difficult to differentiate between these cases based on the symptoms recorded in the questionnaires.

**Figure 3.16 Number of symptoms for hospitalised and not hospitalised cases from the case-control study.**



## 3.5 Completion of the reporting pyramid

Here, the completion of the under-reporting steps in the reporting pyramid (Figure 3.17) are provided.

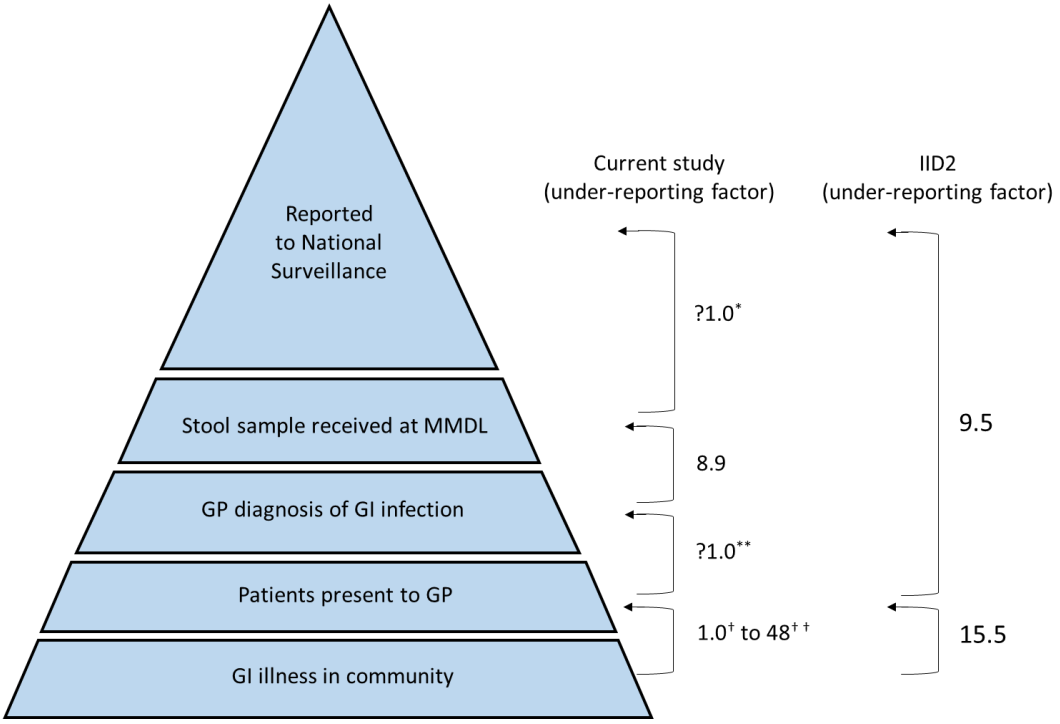
Section 3.2 determines the under-reporting factor between "GI illness in the community" and "patients presenting to the GP" which ranges between 1.0 to 48. This broad range is consistent with the factor of 15.5 that was obtained in the IID2 study (C. Tam, Viviani et al. 2011). It is not possible to have a direct comparison, as the current study did not investigate actual diarrhoeal episodes in the community as was done in IID2. However, the wide range of between 1.0 and 48 does suggest that the actual under-reporting rate will be somewhere between these two figures.

The under-reporting factor between "Patients presenting to GP" and "GP diagnosis of GI infection" is estimated, based purely on the judgement of the authors, to be approximately 1.0 based on the data produced in this study.

Section 3.3 determines the under-reporting factor between "GP diagnosis of GI infection" and stool sample received by MMDL as 8.9 in this study (see Table 3.5 and Figure 3.17). The final under-reporting step between the MMDL and "Reported to National Surveillance" is assumed to be one (or fairly close to it) as this relies on the efficiency of obtaining the pathogen from the sample and inserting and transferring the result between electronic databases. Hence the overall under-reporting factor from "Patients presenting to the GP" to the top of the reporting pyramid is 8.9 =(1.0x1.0x8.9x1.0) which agrees closely with the 9.5 obtained in the IID2 study (see Figure 3.17).

The overall underreporting factor (15.5x9.5=147) from community to national surveillance obtained in the IID2 study stands within the extreme underreporting values (8.9 to 427) obtained in the current study which encompass a very wide range.

**Figure 3.17 Reporting pyramid showing under-reporting ratios for the Infectious Intestinal Disease (IID2) study and the current study.**




* , ** - under-reporting assumed to be close to 1, in the current study, but is unknown, see section 3.5.
†For the most severe symptom (blood in stools) lasting for 14 days or more.
††For diarrhoea or loose stools lasting for one day (this is the most mild symptom and shortest duration).

## 3.6 Discussion

The reporting pyramid for GI infections obtained here is a refinement of that published in the IID2 study (C. Tam, Viviani et al. 2011). For example, in the present study presenting to GP includes a GI diagnosis step and under reporting from community to GP depends on the type and severity of symptoms. Also, the pyramid has an extra level which accounts for the stool samples received at MMDLs. However, where the IID2 study and the current study are comparable the results are broadly similar.

It is worth noting that it is possible to have a GI infection but have no symptoms. This is certainly the case for campylobacteriosis where seroepidemiological studies have shown frequent exposure to *Campylobacter* in humans resulting in a serological response but usually no illness (Teunis, Falkenhorst et al. 2013). However in the current study these asymptomatic cases were not considered because they can only be detected by serology which is not routinely performed in Scotland.

The effect of deprivation was considered in the current study but was not explicitly included in IID2. Here it was found at the community level that there was no difference in the likelihood of making a GP appointment based on the

duration and number of symptoms between individuals belonging to SIMD1 and SIMD5 quintiles. The only apparent difference is found for nausea or vomiting of 14 days duration (Table 3.1). Here, respondents from the least deprived population are more likely to make a GP appointment. Although this is only a small percentage difference (92% compared with 85%) it may contribute to a bias in the reporting rate.

At the GP level it was found that GPs considered recent foreign travel as an important factor in deciding when to request a stool sample, which will be looked at further in chapters 7 and 8. Also, at the GP level there was a higher incidence of RCG3 diagnoses in the most deprived SIMD1 (218.3 diagnoses/100,000) compared with least deprived SIMD5 (157.5 diagnoses/100,000) individuals. This is opposite to what is found for reported campylobacteriosis cases. There was also a higher incidence of MMDL stool samples from SIMD1 compared with SIMD5, but this was not significant, possibly due to small numbers (22 for SIMD1 and 17 for SIMD5). It is worth noting that RCG3 diagnosis comprises all 'Gastroenteritis of possible infectious origin'. Ideally it would have been useful to look at higher resolution (specifically for *Campylobacter*) at the GP level but this was not practical with the datasets that were available in the current study.

At the reported case level, data were not available on the proportion of RCG3 diagnosed cases involving hospitalisation. However, considering only campylobacteriosis, the case-control study found proportionally more case patients resident in SIMD1 data zones were hospitalised compared with SIMD5.

When comparing the frequency, duration and number of symptoms between hospitalised and not-hospitalised campylobacteriosis cases there were few differences, however diarrhoea or loose stools were more common in not-hospitalised (98%) compared with hospitalised (92%) case patients. This finding may appear surprising because hospitalised cases might be expected to include the most symptomatic patients. However, blood in stools (Figure 3.16), a more severe symptom, is indeed more common (though not statistically significantly so) in hospitalised cases (34% compared with 27%).

A weakness in the reporting pyramid was that it did not consider cases in the community. Having an estimate of the number of IID, or even better campylobacteriosis cases in the community would have helped understanding any biases at this step of the pyramid. Instead, information from controls was used to determine whether an individual was likely to attend a GP based on the type and duration of symptoms that they might suffer during an episode of IID.

## 3.7 Conclusions

It was possible to generate a reporting pyramid in Scotland. The under-reporting rates were similar to those observed from the IID2 study

Socioeconomics did not appear to be important in a number of steps of the reporting pyramid. However the following differences were observed:

- GPs indicated that recent foreign travel was a very important consideration when requesting a stool sample from an individual with presumptive IID and for cases deciding to submit a stool sample.

- There was also some evidence that having prolonged "nausea or vomiting" symptoms was more likely to lead to an individual from a least deprived background (SIMD5) making a doctor's appointment.
- The incidence of diagnoses by GPs of IID (RCG3) was higher for SIMD1 compared with SIMD5
- Hospitalisation was proportionally higher in campylobacteriosis cases from SIMD1 compared with SIMD5

Hospitalised and not-hospitalised campylobacteriosis cases did not differ by frequency, duration and number of symptoms.

# 4. Reported Case Study

## 4.1 Introduction

This chapter provides the analysis of human campylobacteriosis case data from Scotland. It addresses Objective 3 'Collect reported case data', and Objective 6 'Analyse reported case data'.

Throughout, comparisons are made with a previous Food Standards Agency Scotland funded project (S14004) on "Factors associated with geographical and temporal variation in campylobacteriosis in humans." This project studied campylobacteriosis in Scotland between 2000 – 2006. For brevity this will be referred to as "the geography study" (Anon. 2007). This study found:

- that the incidence of *Campylobacter* infection varies considerably from region to region, in particular some health boards reported more cases than others.
- differences in the geographic distribution of *Campylobacter* infections within Scotland caused by differences in exposure to infection. Deprivation was found to be a protective factor, with higher rates of *Campylobacter* infection reported in less deprived areas, a feature that was attributable to reduced overseas travel. At least part of the difference is likely to be a result of real differences in rates of infection, although some may be due to differences in ascertainment.
- in combination with findings from an FSAS funded project on source attribution of *Campylobacter* infection (project S14006), that retail chicken as well as ruminants are important sources of human campylobacteriosis.

This chapter updates the descriptive and epidemiological analysis with a particular focus on deprivation. In particular:

- to determine whether there are still proportionally fewer cases in deprived areas compared with less deprived areas;
- to describe the epidemiology of reported cases of campylobacteriosis across Scotland.

To achieve this the following three approaches were taken:

1) Investigation of risk factors and the pattern of disease were identified using both descriptive and analytical (i.e. Poisson, logistic and multinomial regression) epidemiology. (Section 4.3);
2) The effect of proximity to a GP practice on likelihood of reporting was assessed by combining reported case data with the locations of GP practices (Section 4.4) and
3) Changes in secular trends (i.e. trends over a long period) identified from long term reported campylobacteriosis data (1990-2017) (Section 4.5).

## 4.2 Overview of data sources

Health Protection Scotland (HPS) collates laboratory confirmed reports of human campylobacteriosis from each of the health boards on its ECOSS (The Electronic Communication of Surveillance in Scotland) database. The following data were obtained from HPS:

**Collect retrospective case data.** Non-identifiable reported case data (age, date of report, gender, health board and data zone) were obtained for the four years and three months previous to the start of the current study (1st January 2012 to 31st March 2016) across Scotland. These comprised 26,374 cases.

**Collect prospective case data.** Non-identifiable reported case data (age, date of report, gender, health board and data zone) were obtained for the two years from 1st April 2016 to end of March 2018. This comprised 11,236 cases. Since the case-control study ran for an additional 5 months, summary ECOSS data were obtained from HPS providing the number of reported cases by month by health board. These comprised 3215 further cases.

**Long term summary epidemiological case data.** Human summary campylobacteriosis case data from Scotland during 1990 to 2011 (n=112,230) were obtained from the literature (N. J. C. Strachan, Rotariu et al. 2013). This combined with the above data enabled long term trends to be determined.

## 4.3 Perform descriptive and analytical epidemiology on retrospective and prospective campylobacteriosis cases

### 4.3.1 Aims

This section aims (i) to provide a description of human campylobacteriosis in Scotland stratified by age, gender, deprivation, health board, rurality and temporal (i.e. changes over time) trends and (ii) to identify risk factors for human campylobacteriosis and in particular those factors that may be associated with and differentiate between populations based on deprivation.

### 4.3.2 Materials and Methods

#### 4.3.2.1 Data

Collection of reported case data from 1st January 2012 to 31st March 2018 is described in 4.2.

Apart from the case data that was analysed there were the following non-disease data used in the analysis:

Scottish Index of Multiple Deprivation (SIMD2012): this comprised 6505 data zones from across Scotland which had SIMD scores between 0.94 (least deprived) and 89.89 (most deprived) (http://www.gov.scot/Topics/Statistics/SIMD/DataAnalysis/Background-Data-2012. Each data zone was allocated to an SIMD quintile where SIMD1 is the most deprived and SIMD5 is the least deprived. The latitude and longitude of the

centre of each data zone was obtained, as well as the population and also a shape file to enable plotting on a map. Throughout this study the data from SIMD2012 were used rather than the recent release from 2016. This was because the overall deprivation score was available for each data zone and used in the Poisson regression as a continuous variable. (For information, Figure A4.1.7 in Annex 4.1 shows that there is a strong correlation (87%) between SIMD2012 and SIMD2016). Further, the Scottish government uses a number of indicators to monitor poverty in the Scottish population. The majority of these were relatively flat over the time period where data were available (https://nationalperformance.gov.scot/measuring-progress/national-indicator-performance).

Human population data: The number of people at the mid-point of each year from 2012 to 2017 and stratified by five year age groups (0-4, 5-9, etc.) and by health board were made using the National Records of Scotland (NRS) (https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates). Summary table and maps are provided in Annex 4.1 (Table A4.1.3 and Figures A4.1.3, A4.1.4 and A4.1.5 and A4.1.6).

Private water supplies: the numbers of properties on private water supplies (PWS), including postcode, were obtained from local authorities across Scotland. Summary table and maps are provided in Annex 4.1 (Table A4.1.1 and Figure A4.1.1).

Farm animal numbers: these were obtained from the 2012 Agricultural census (http://agcensus.edina.ac.uk/ ). This comprised Cattle, Pigs, Broilers, Ducks, Geese, Poultry, Sheep, Horses and Deer at a spatial resolution of 2x2 $km^2$. These were aggregated into SIMD data zones. Summary table and maps are provided in Annex 4.1 (Table A4.1.2 and Figure A4.1.2).

#### 4.3.2.2 Descriptive epidemiology

Graphs and tables were generated to illustrate how human campylobacteriosis varied by age, gender, deprivation, health board, rurality and time (annual, season or monthly). Confidence intervals (95% CI) for campylobacteriosis incidence were calculated by finding the standard deviation over the number of years being considered and assuming a normal distribution (Caulcutt 1983). P values quoted were calculated by student's t-test unless stated otherwise.

SPSS Statistics v24 was utilised to determine the difference in incidence between health boards by Analysis of Variance using Tukey's honest significant difference (Tukey 1949) with post-hoc correction (Bonferroni) for multiple comparisons. The student's t-test was used to compare average incidence values between groups (Clifford-Blair, Higgins 1980).

To visualise the number of reported cases and disease incidence of human campylobacteriosis maps were produced in ArcMap 10.5 (http://www.arcgis.com).

### 4.3.2.3 Univariate and multivariate Poisson regression

Univariate and multivariate Poisson regression (Gardner, Mulvey et al. 1995, Osgood 2000) analysis was performed on the case data (1st January 2012 to 31st March 2018), using SPSS Statistics v24. Briefly, the outcome (number of cases in data zone *i* - $N_{cases_i}$) was fitted using a Poisson distribution offset by the natural logarithm of the population ($N_{people_i}$) for each SIMD data zone, *i*. Thus, the model takes the form:

$$N_{cases_i} \sim Poisson(\lambda_i) \quad (4.1)$$

$$Ln(\lambda_i) = \beta_j X_{ij} + Ln\left(N_{people_i}\right) + a_0 \quad (4.2)$$

where, $\lambda_i$ represents the mean and variance of the number of cases in each datazone *i*, $X_{ij}$ is the matrix of risk factors denoted by *j,* in each data zone *i*, $\beta_j$ are regression coefficients and $a_0$ is the intercept. The list of risk factors used in the univariate Poisson regression analysis were:

- position of the data zone (latitude and longitude)
- SIMD score (low means least deprived population and high most deprived population)
- human density (people/km$^2$)
- density of properties on private water supplies (properties /number of people)
- poultry density (poultry/km$^2$)
- cattle density (cattle/km$^2$)
- sheep density (sheep/km$^2$).

All factors having a P-value <0.25 were introduced in the multivariate Poisson regression analysis. A p-value of 0.25 was selected as this is a relaxed value of p and more stringent setting of p to <0.05 can fail in inclusion of variables known to be important (Bursac, Gauss et al. 2008)

### 4.3.2.4 Univariate and multivariate binary logistic regression

Univariate and multivariate binary logistic regression analysis (Kleinbaum, Klein 2010, Cox 1958) looked for differences between risk factors for cases from the *most* and *least* deprived quintiles. Here *"controls"* (0s) were defined as cases from the most deprived data zones (SIMD1 quintile) and *"cases"* (1s) were cases from the least deprived data zones (SIMD5 quintile), respectively. The univariate logistic model fits a logit function to the risk factors,

$$Ln\left(\frac{p_i}{1-p_i}\right) = \beta_j X_{ij} + b_0 \quad (4.3)$$

where $p_i$ is the probability of an individual *i* to be a "case", given the *j*'th risk factor value $X_{ij}$ (e.g. for gender: male or female.) for that individual. $\beta_j$ and $b_0$ are the slope and constant of the regression. The following risk factors were used in the univariate binary-logistic regression: health board, age group, time of year, gender, latitude, longitude, density of a specific animal group (e.g. cattle, sheep, and poultry), density of human population, and presence/absence of properties on PWS in data zone. The *odds* (or the ratio of *"cases"/"controls"* - here the ratio of *"least deprived cases"/"most deprived cases"*) was used to

calculate the *odds ratio* ($OR_j$) for the *j*'th risk factor (e.g. female), compared with the reference ($ref$) (e.g. male):

$$OR_j = \frac{\left(\frac{least\ deprived\ cases}{most\ deprived\ cases}\right)_j}{\left(\frac{least\ deprived\ cases}{most\ deprived\ cases}\right)_{ref}} \quad (4.4)$$

The relationship between the regression coefficients $\beta_j$ and $OR_j$ is

$$OR_j = e^{\beta_j} \quad (4.5)$$

All factors from the univariate logistic regression analysis having a P-value <0.25 were introduced into the multivariate analysis in one step using equation (4.3) (https://www.ibm.com/support/knowledgecenter/en/SSLVMB_24.0.0/spss/regression/logistic_regression_methods.html ).

The risk factors were split as follows (with arbitrarily selected ref points):

- health board (11 health board areas; - Tayside – with the highest mainland campylobacteriosis incidence was chosen as *ref*)
- age group (0-4, 5-24, 25-64 and 65+ years old (*ref*))
- time of year ("Summer" as (May, June, July, August) and "Rest of year" *(ref*))
- gender (female and male (*ref*))
- latitude (continuous variable), longitude (continuous variable)
- human population density ((<200/km$^2$ – rural, 200-2500/km$^2$ – peri-urban, and ≥2500/km$^2$ – urban (*ref*))
- private water supplies (present, absent(*ref*) – this treated as categorical variable because of large number of data zones with none present)
- cattle density ("Cattle density-Low" (0-9.03 cattle/km$^2$); "Cattle density-Mid1" (9.04-23.7 cattle/km$^2$); "Cattle density-Mid2" (23.8-44.4 cattle/km$^2$) and "Cattle density-High" (44.5-216.8 cattle/km$^2$) (*ref*))
- sheep density ("Sheep density-Low" (0-9.7 sheep/km$^2$); "Sheep density-Mid1" (9.8-38.6 sheep/km$^2$); "Sheep density-Mid2" (38.7-78.4 sheep/km$^2$) and "Sheep density-High" (78.5-492.8 sheep/km$^2$) (*ref*))
- poultry density ("Poultry density-Low" (0-1.24 poultry/km$^2$), "Poultry density-Mid1" (1.25-10.62 poultry/km$^2$), "Poultry density-Mid2" (10.63-222.5 poultry/km$^2$) and "Poultry density-High" (222.6-19602 poultry/km$^2$)(*ref*)).

#### 4.3.2.5 Univariate and multivariate multinomial logistic regression

Univariate and multivariate multinomial logistic regression (Varga, Middleton et al. 2012) was used since it utilises all 5 deprivation quintiles (Note the binary logistic regression compares only quintiles 1 and 5 but in an identical manner). All 5 SIMD deprivation quintiles are categories of the outcome variable, where SIMD1 cases (most deprived) are *"controls"* (0s) and SIMD2 to SIMD5 cases (less deprived) are *"cases"* (1, 2, 3 or 4). Since there are 4 *"case"* groups, there are four logit functions similar to eq. (4.3) used to determine the regression coefficients. Four odds ratios are then calculated (equations (4.6) to (4.9) for each risk factor as in eq. (4.4), the only difference being that *"least deprived cases"* are replaced by one of the *"less deprived case"* categories (i.e. cases in

SIMD2 or SIMD3 or SIMD4 or SIMD5). The *"most deprived cases"* category corresponds to cases in SIMD1 quintile. Hence for the *j'*th risk factor the corresponding four odds ratios are:

$$OR_{j,SIMD2\ v1} = \frac{\left(\frac{SIMD2\ cases}{SIMD1\ cases}\right)_j}{\left(\frac{SIMD2\ cases}{SIMD1\ cases}\right)_{ref}} \quad (4.6)$$

$$OR_{j,SIMD3\ v1} = \frac{\left(\frac{SIMD3\ cases}{SIMD1\ cases}\right)_j}{\left(\frac{SIMD3\ cases}{SIMD1\ cases}\right)_{ref}} \quad (4.7)$$

$$OR_{j,SIMD4\ v1} = \frac{\left(\frac{SIMD4\ cases}{SIMD1\ cases}\right)_j}{\left(\frac{SIMD4\ cases}{SIMD1\ cases}\right)_{ref}} \quad (4.8)$$

$$OR_{j,SIMD5\ v1} = \frac{\left(\frac{SIMD5\ cases}{SIMD1\ cases}\right)_j}{\left(\frac{SIMD5\ cases}{SIMD1\ cases}\right)_{ref}}. \quad (4.9)$$

where when considering the risk factor "gender", *j* can be "female" and *ref* "male".

All risk factors with P<0.25 from the univariate analysis were introduced in the multivariate analysis simultaneously. Then non-significant factors were removed stepwise from the analysis, until only those with P<0.05 were left and this comprised the final model.

The risk factors used in the analysis were the same as in 4.3.2.3.

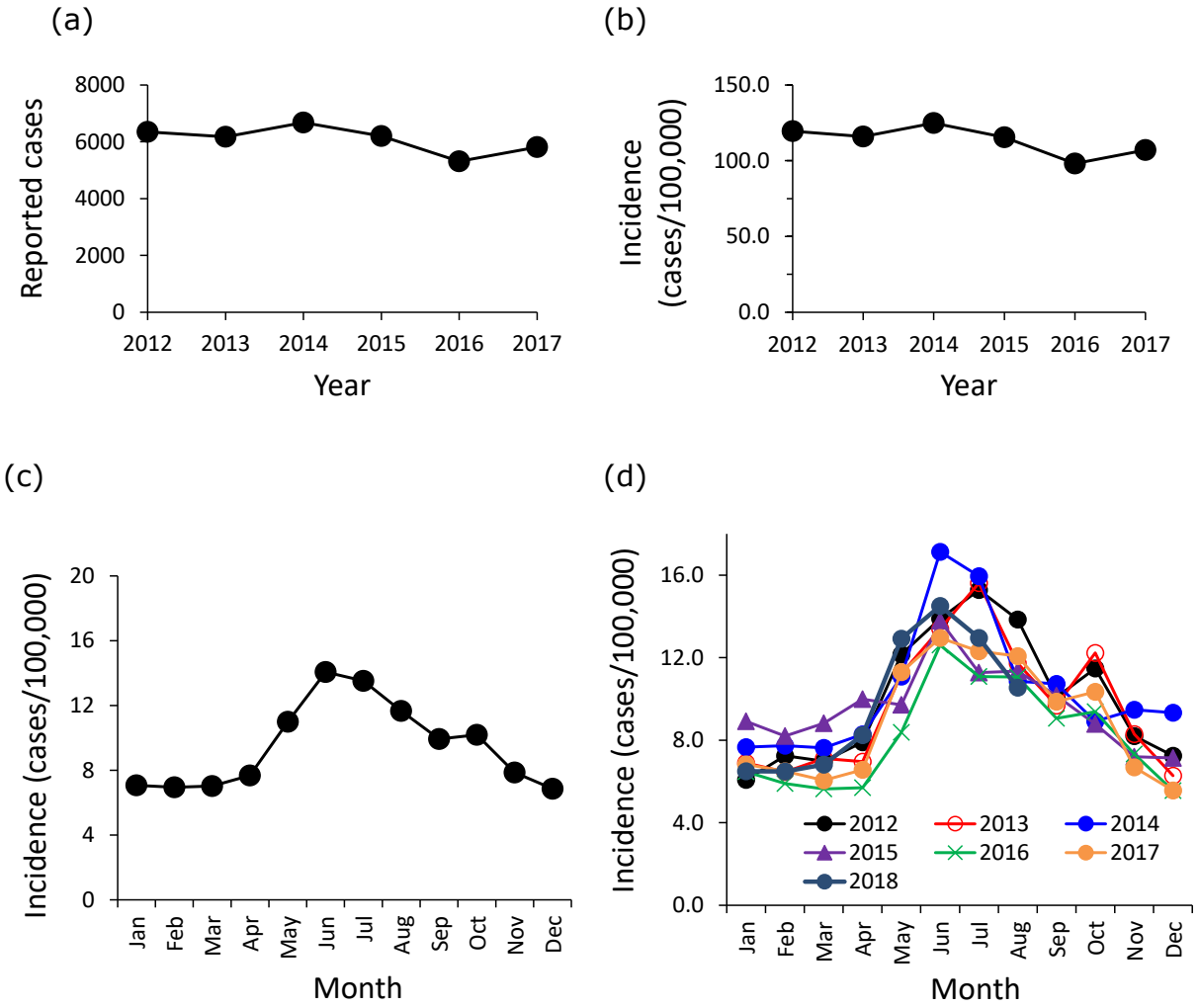
## 4.3.3 Results and discussion

### 4.3.3.1 Human campylobacteriosis incidence rate in Scotland, January 2012 to March 2018.

In Scotland the annual average number of reported campylobacteriosis cases between 2012 and 2017 was 6087. Incidence declined slightly during 2015 and 2016, then increased in 2017 (Figure 4.1(a) and (b)). There is a consistent summer (May to August) peak (Figure 4.1(c)). The summer incidence of 13 cases/100,000/month (95% CI 11.3-14.7) was significantly higher (P=0.0028) than for the rest of the year (8.0 cases/100,000/month (95% CI 6.7-9.3)). A second smaller peak occurred in October of 2012 and 2013 and less so of 2017, but not for other years (Figure 4.1(d)). The incidence of human campylobacteriosis during 1st January 2012 to 31st March 2018 (112.1±7.0 cases/100,000) was significantly higher (P = 0.009) than that recorded during 2000-2006 in the geography study (97±9 cases/100,000) (Anon. 2007).

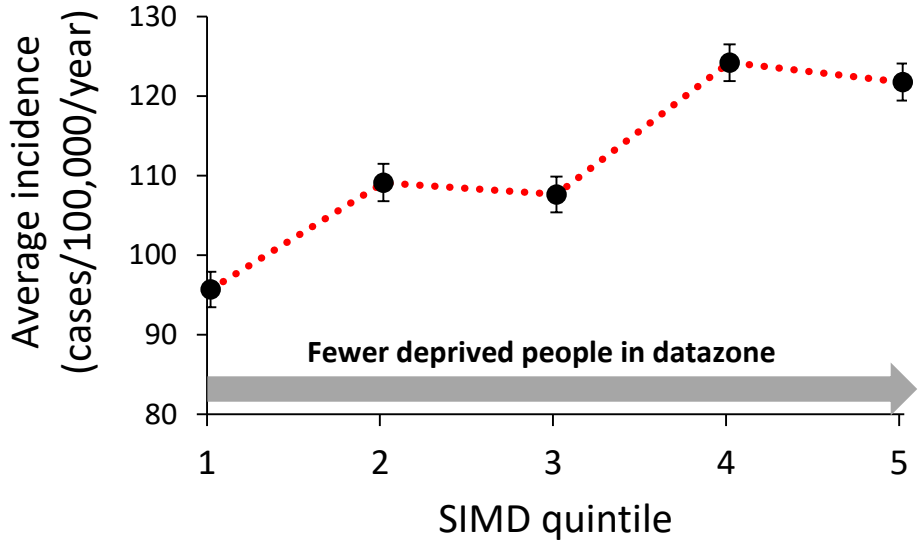
The graphs in Figures 4.1(a)-(d) depict this higher incidence rate and consistent seasonal variation. They do not suggest ongoing increase during 2012-2017 and incidence may therefore have reached a plateau.

**Figure 4.1. Incidence of human campylobacteriosis in Scotland**




(a) Reported cases and (b) incidence of human campylobacteriosis in Scotland Jan 2012 -Dec 2017, (c) monthly incidence in Scotland between 1st Jan 2012 – 31st Aug 2018 and (d) monthly incidence by year 1st Jan 2012 – 31st Aug 2018.

**Figure 4.2. Variation in incidence by SIMD Quintile (2012-2017)**

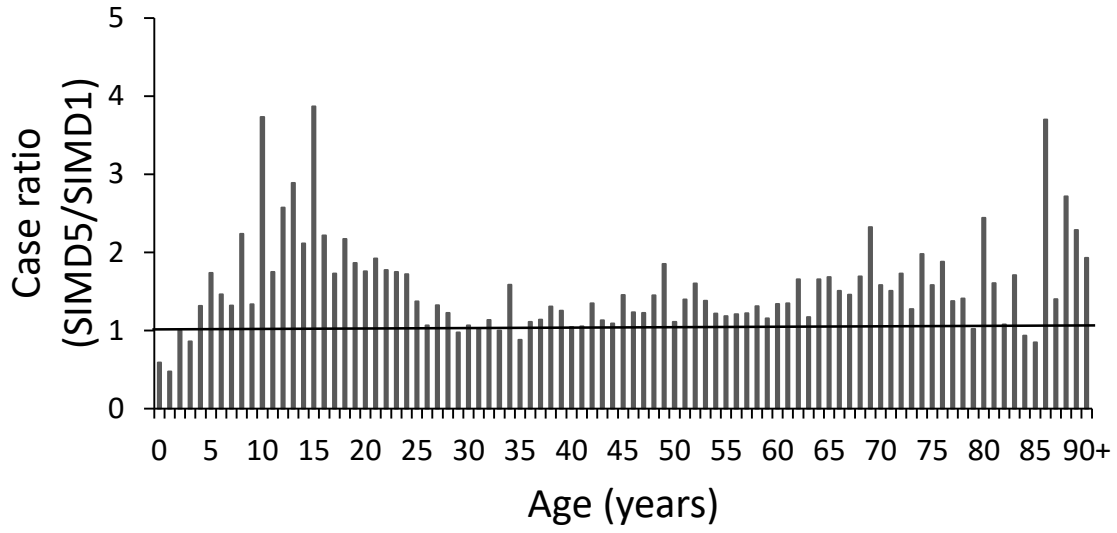



The largest number of deprived people live in SIMD quintile 1 whilst the fewest live in SIMD quintile 5.

There was an excess of cases (19%) in the four less deprived SIMD quintiles compared with the most deprived SIMD quintile (Figure 4.2). During the previous geography study (2000-2006) the 26% excess of cases was comparable (Anon. 2007). Although the percentage has reduced between the two studies it still corresponds to a substantial proportion of cases. Also, in the present study there was a significantly ($P=7.0x10^{-6}$) higher number of cases in less deprived areas than most deprived, and the difference can be observed across most ages (Figure 4.3). However, it appears that there is an excess of cases in the most deprived population for young children (e.g. <5 years old). A similar result was reported from Connecticut in the USA (Bemis, Marcus et al. 2014).

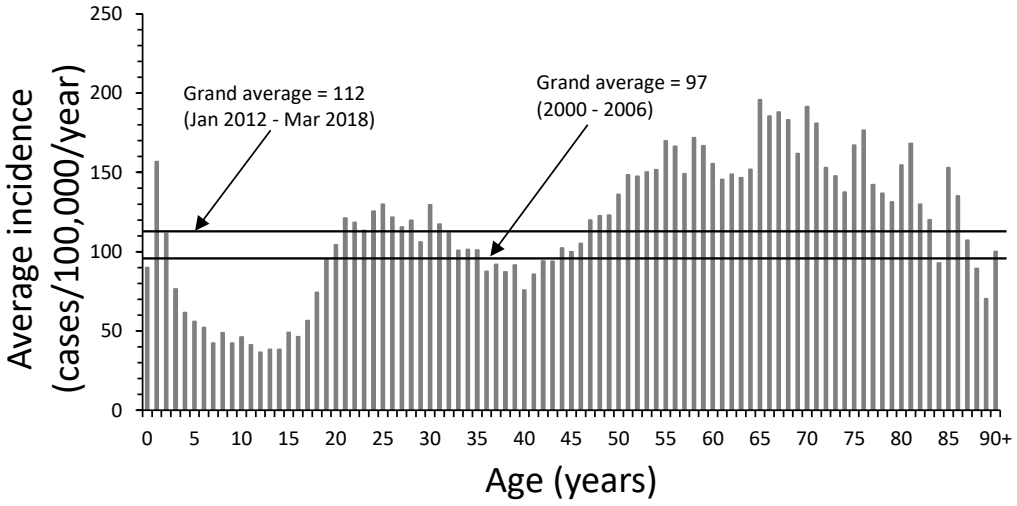
Annex 4.1, Figure 4.1.6 shows that there is a higher proportion of people in Scotland in the older age groups (>39 years) that are in the least deprived SIMD5 quintile.

**Figure 4.3. Ratio of cases resident in least deprived (SIMD5) to number in most deprived (SIMD1) areas / by age.**




SIMD5 is the least deprived quintile. The horizontal line represents an equal ratio between least to most deprived areas, thus bars above the line are age classes with a higher proportion of cases residing in least deprived areas.

**Figure 4.4. The average incidence of campylobacteriosis cases in Scotland by age**




1st Jan 2012 -31st Mar 2018. The horizontal line (grand average) represents the over-all-ages average incidence for two time periods.

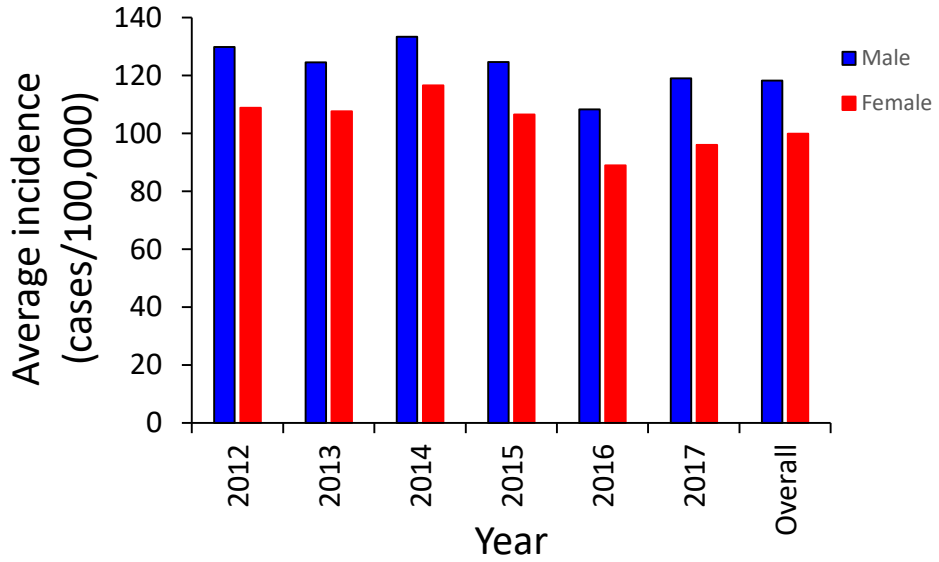
The average incidence across the 6.25 year period of this study was stratified by age (Figure 4.4).

The incidence rises from infancy to 1 year olds and then falls, and remains low, to age 15 years before rising again in young adults. This was also found in the previous geography study and is widely reported elsewhere (e.g. for England and Wales (Gillespie, O'Brien et al. 2008)).

The incidence among Scots aged over 50 years (148.9±8.6 cases/100,000/year) was significantly (P<0.05) higher than the average for all ages (112.1±7 cases/100,000/year) (Figure 4.4). This difference appears much greater than the previous geography study [see Figure 5.2 of that study] (Anon. 2007).

Figure 4.5 shows the annual incidence of campylobacteriosis cases by gender. The incidence in the male population (118/100,000) was significantly (P=0.0008, by two sample t-test) higher than the incidence in female population (100/100,000). Overall, there was an excess (approx. 11%) of male cases as has been reported in the previous geography study (12% excess) and in the literature (Gillespie, O'Brien et al. 2008). Evidence has also been presented that physiological factors rather than behavioural differences may be the dominant explanation for the difference in very young children (N. J. Strachan, Watson et al. 2008).

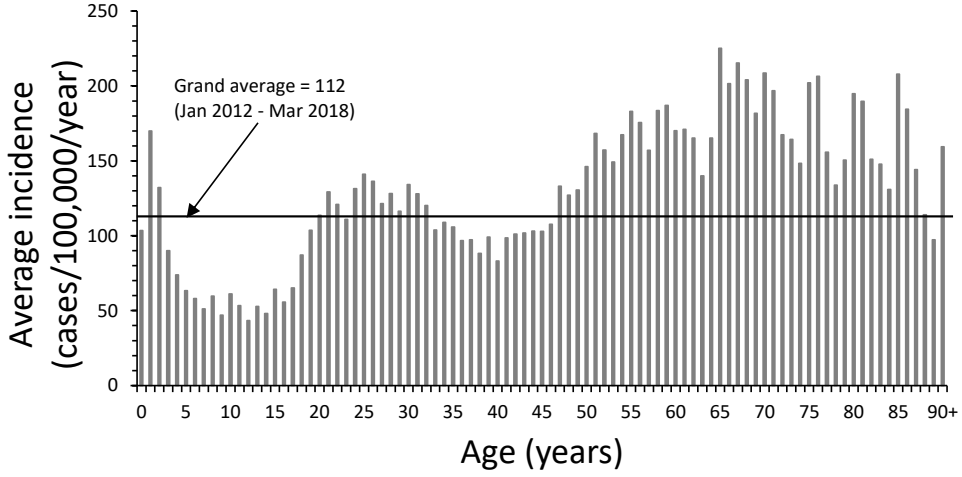
**Figure 4.5. The average incidence of campylobacteriosis cases in Scotland by gender and year (1st Jan 2012- 31st March 2018).**



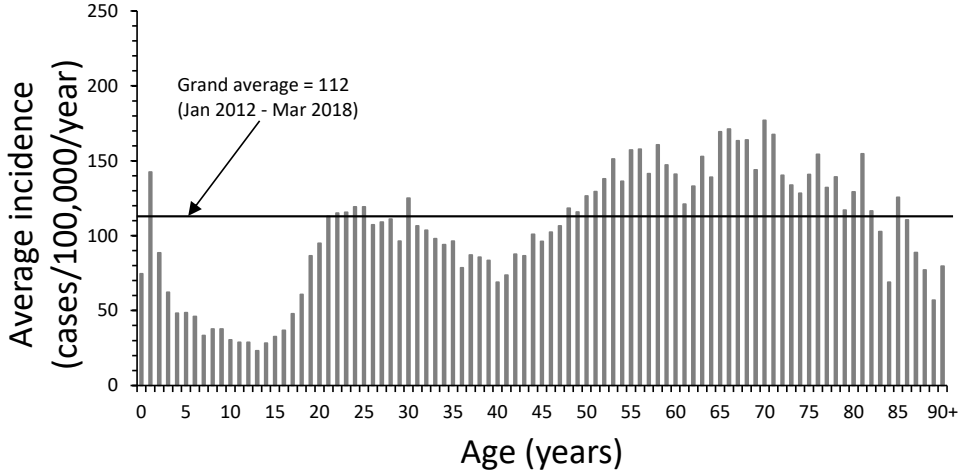

1st Jan 2012 -31st Mar 2018.

**Figure 4.6. The average incidence by gender stratified by age**

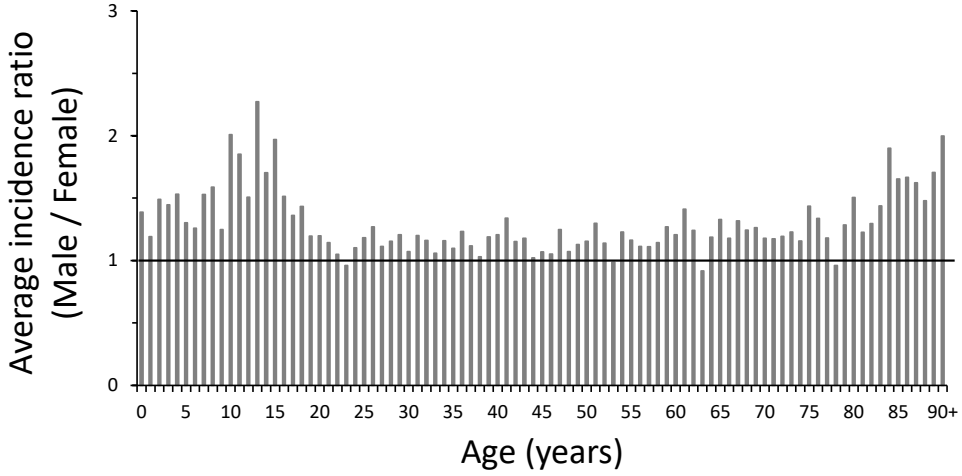
(a)




(b)



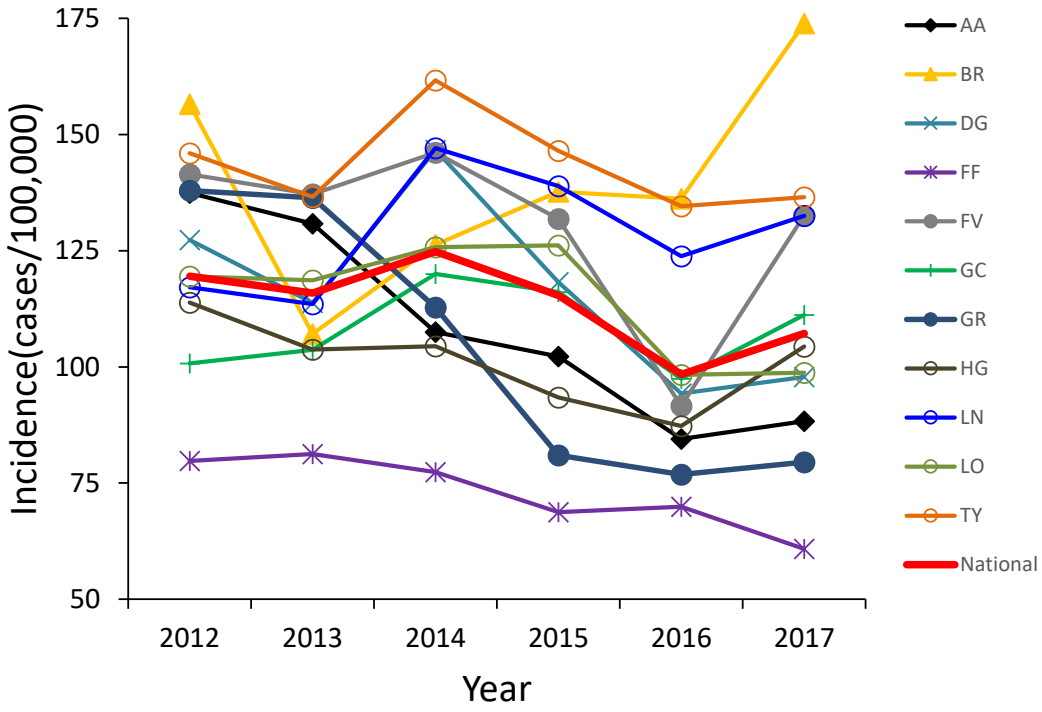

(c)




(a) male, (b) female and the (c) male:female incidence ratio stratified by age. The horizontal line in (a) and (b) (grand average) represents the over-all-ages average incidence.

When the average incidence of cases was stratified by age for each gender (Figure 4.6), it was shown that overall the incidence across ages was, on average, higher in males than in females (incidence ratio >1, P=0.0001). This is different than in the previous geography study, where males had higher incidence rates than females only for <18 years of age and for elderly people (>65 years). It is unclear why there should be this change but it is worth monitoring to establish whether this trend continues.

**Figure 4.7. Incidence of human campylobacteriosis stratified by year for mainland health boards in Scotland.**




---

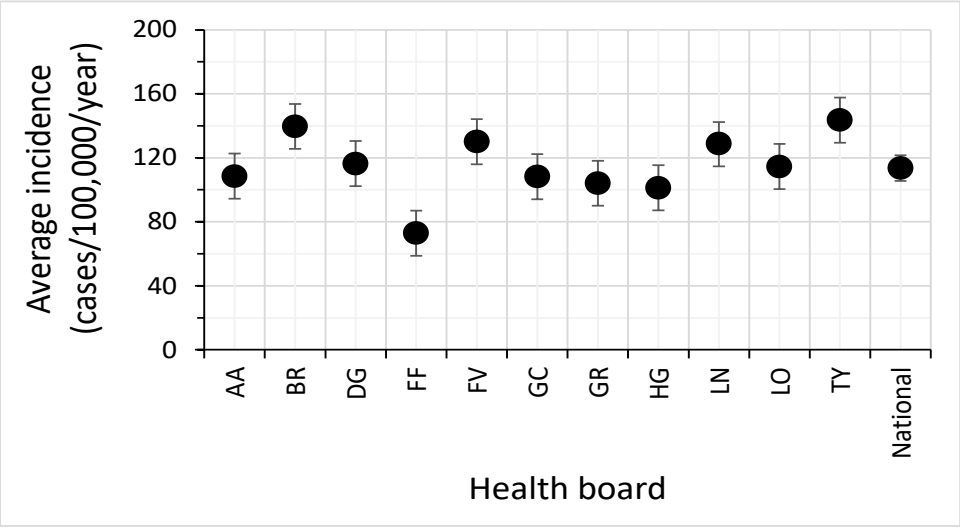
The human campylobacteriosis incidence was highest for Tayside (TY) health board (143.6 cases/100,000 people) (Table 4.1 and Figure 4.8) and lowest for Fife (FF) (73 cases/100,000 people). The incidence in AA, FF and GR appears to have a decreasing trend during the study period (Figure 4.7). The analysis of variance performed to determine whether there are differences in incidence between each of the mainland health boards shows that the incidence in FF was significantly lower (P<0.05) than AA, BR, FV, GC, LN, LO and TY (see Table 4.2 and Figure 4.8).

**Table 4.1. Campylobacteriosis incidence by health board.**

| Health board | Average incidence (cases/100,000/year) (Jan 2012 – Mar 2018) |
|---|---|
| Ayrshire & Arran (AA) | 108.4 |
| Borders (BR) | 139.6 |
| Dumfries & Galloway (DG) | 116.4 |
| Fife (FF) | 73.0 |
| Forth Valley (FV) | 130.1 |
| Greater Glasgow & Clyde (GC) | 108.2 |
| Grampian (GR) | 104.1 |
| Highland (HG) | 101.1 |
| Lanarkshire (LN) | 128.8 |
| Lothian (LO) | 114.5 |
| Tayside (TY) | 143.6 |
| Orkney (OR) | 191.4 |
| Shetland (SH)* | 123.2 |
| Western Isles (WI) | 80.3 |
| National | 112.1 |

* Jan to Aug 2012 - no data

**Figure 4.8. The average incidence of campylobacteriosis infections by health board**




Error bars represent 95% CIs

**Table 4.2. Difference in incidence between health boards by Analysis of Variance**

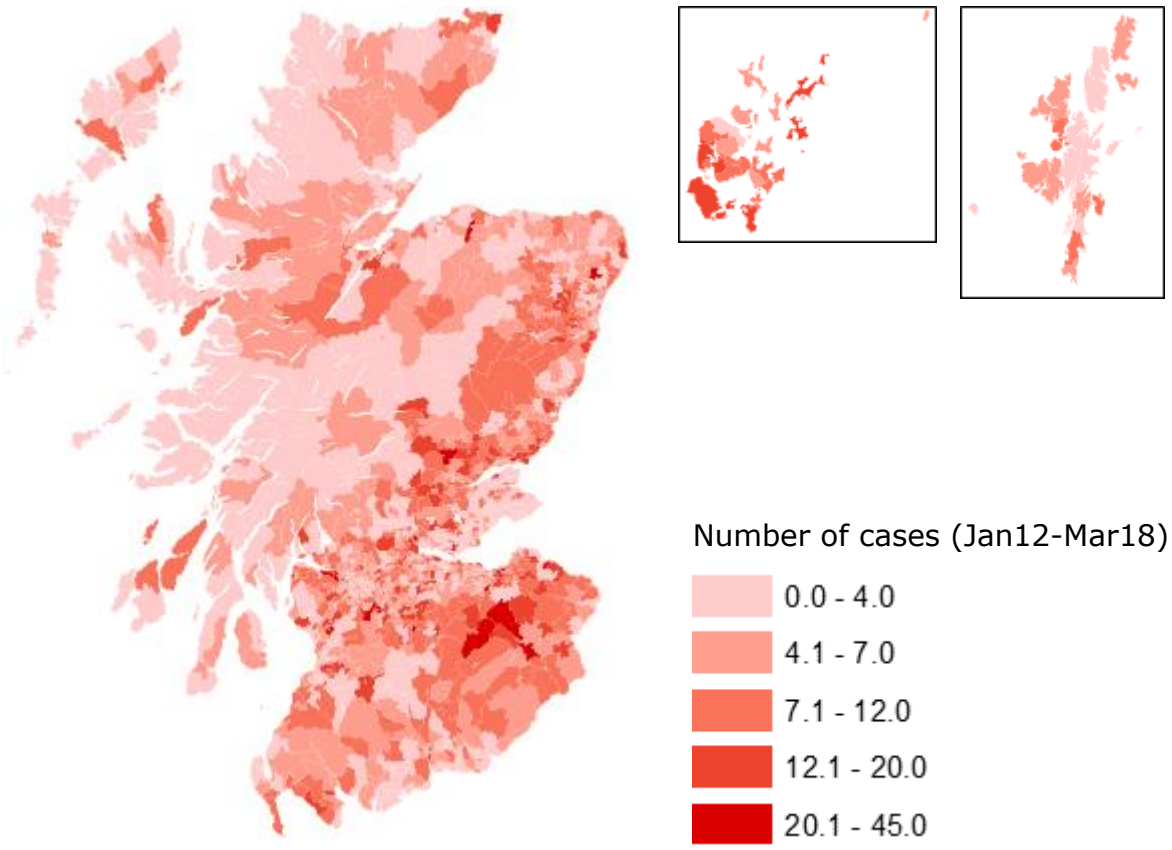
| Contrast | Difference in incidence between health boards (cases/100,000/ year) | P-value (of seeing observed difference or greater) |
|---|---|---|
| AA>FF | 35.6 | 0.028 |
| AA<TY | -35.2 | 0.031 |
| BR>FF | 66.7 | <0.001 |
| BR>GR | 35.5 | 0.028 |
| BR>HG | 38.4 | 0.012 |
| DG>FF | 35.5 | 0.003 |
| FF<FV | -57.22 | <0.001 |
| FF<LN | -55.9 | <0.001 |
| FF<LO | -41.6 | 0.005 |
| FF<GC | -35.3 | 0.030 |
| FF<TY | -41.6 | <0.001 |
| GC<TY | -35.4 | 0.029 |
| GR<TY | -39.6 | 0.009 |
| HG<TY | -42.5 | 0.004 |

Analysis of Variance using Tukey's honest significant difference with post-hoc correction (Bonferroni) for multiple comparisons.

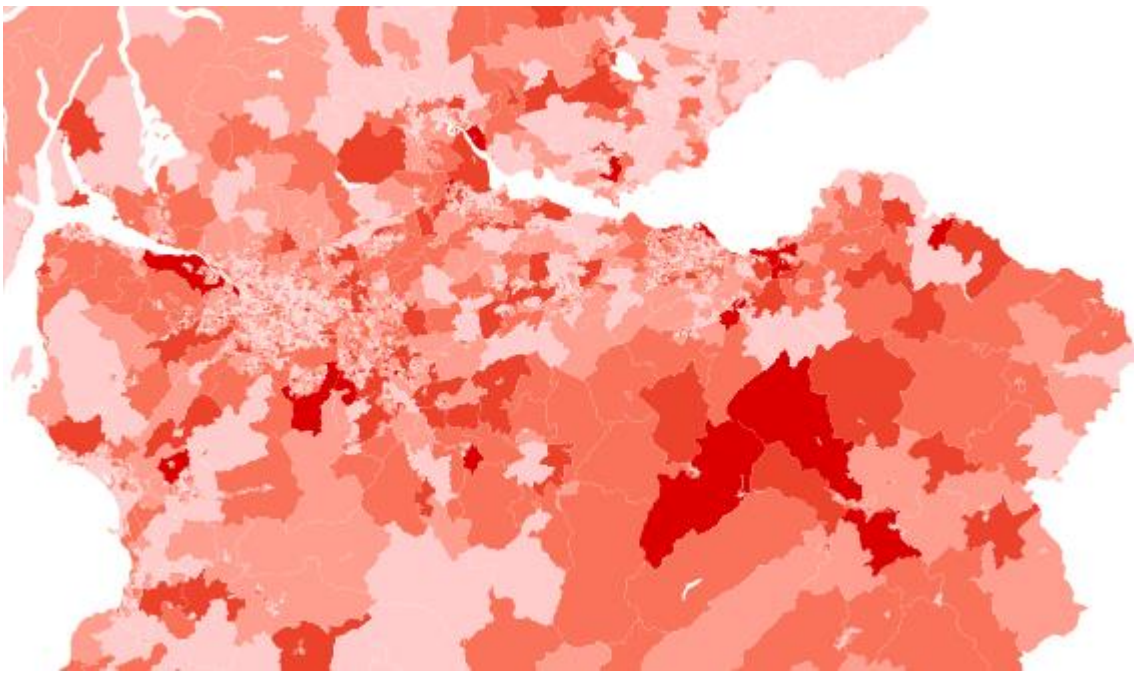
The number of reported cases (Figure 4.9(a) and (b)) appear to be higher in the east compared to the west but elsewhere the data appear to be quite heterogeneous. The incidence in Fife is lower than in other regions (Figure 4.9(c) and (d) and Figure 4.10).

**Figure 4.9. Number and incidences of *Campylobacter* cases in Scotland by SIMD data zones**

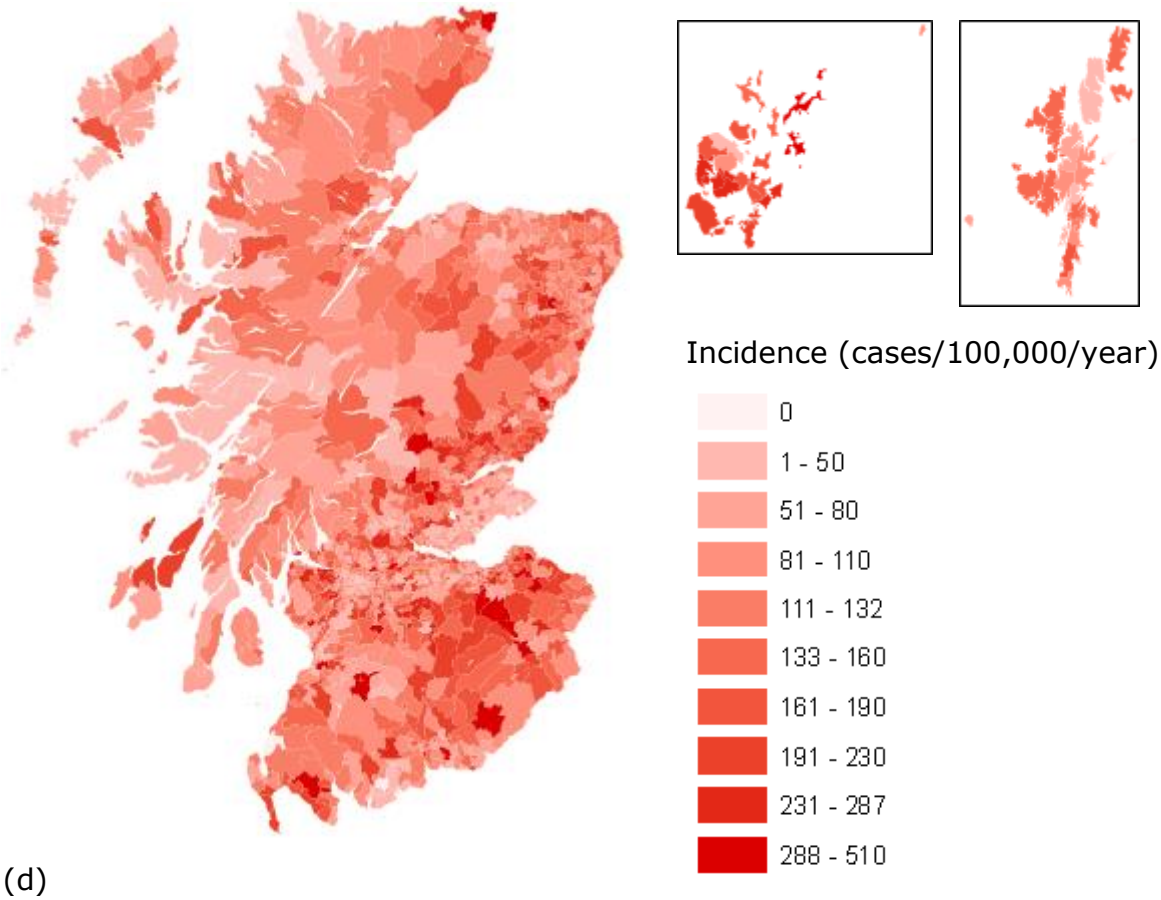
(a)



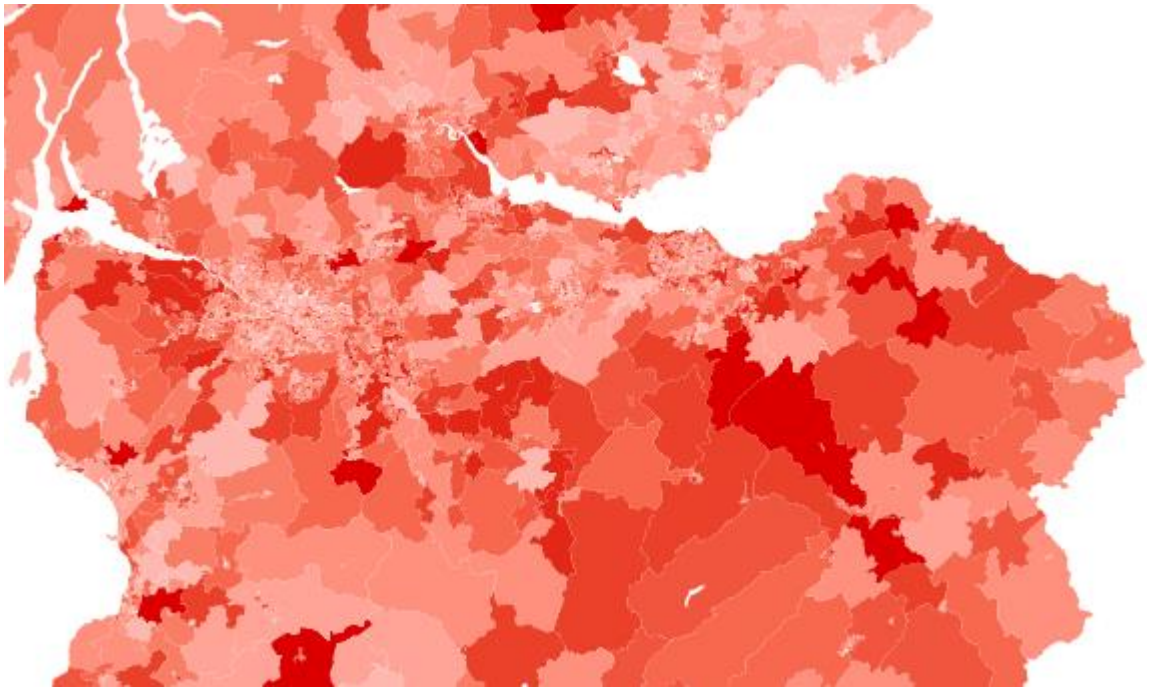

(b)



(c)



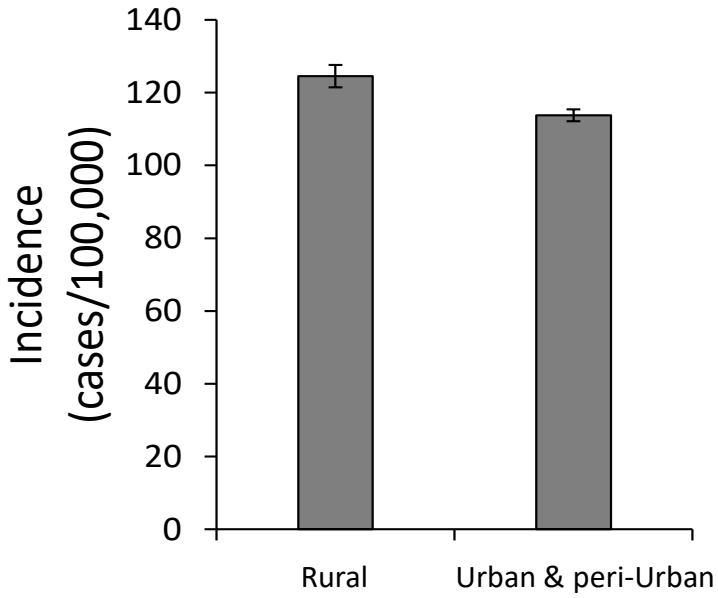

(d)



Numbers of *Campylobacter* cases (a) throughout Scotland and (b) focus on the central belt, and incidence of *Campylobacter* cases per year (c) throughout Scotland and (d) focus on the central belt of human campylobacteriosis for SIMD data zones (1st Jan 2012 –31st Mar 2018).

The incidence in rural regions was significantly higher than urban and peri-urban (Figure 4.10, $P=5.7\times10^{-10}$ by two sample t-test). This difference in incidence is 11.5%. It was not possible for a direct comparison of this result with the finding from the previous geography study (Anon. 2007), as the data on population density was available at different spatial resolutions (postcode sector level compared to data zone level in the present study). However, that study also found incidence higher in rural areas.

**Figure 4.10. Incidence of rural and urban populations for campylobacteriosis cases**




Average incidence and 95% CIs were calculated at data zone level. (Threshold population density: Rural ≤200 people/km$^2$; Urban and peri-Urban >200 people/km$^2$).

#### 4.3.3.2 Risk factors associated with human campylobacteriosis in Scotland. Results from univariate and multivariate Poisson regression

(i) Univariate Poisson regression

The univariate Poisson regression analysis (Table 4.3) shows that increasing cattle and sheep densities were positively associated ($P<0.05$) with increasing disease incidence in this study but were negative (protective) in the previous geography study (2000-2006) (Anon. 2007). It is unclear why this has changed. It is known that sheep and cattle shed *Campylobacter* so it would be expected they would add a potential environmental risk. However, to get opposite results indicates an interaction with some of the other factor(s). Increasing poultry density was associated with decreasing disease incidence ($P<0.05$) as in the previous geography study. This result is perhaps surprising because it can be hypothesised that increasing poultry density would increase the risk of contracting campylobacteriosis from the environment. However, in New Zealand poultry farm distance was also found not to be a risk factor for human campylobacteriosis (Spencer, Marshall et al. 2012).

Increasing human population density (peri-urban/urban) was associated with decreasing disease incidence ($P<0.05$) as in the previous geography study. As deprivation increases then campylobacteriosis incidence decreases ($P<0.05$). Increasing private water supply (PWS) density was not associated with increasing disease incidence whereas in the previous geography study it was positively associated. Note that the current study also contained Highland PWS data which was not included in the previous geography study and may have affected the results. Further it is possible that the quality of private water supplies may have improved because grants have been available from local councils for this purpose. Longitude was positively associated with disease incidence (i.e. higher incidence towards east) as was found in the previous geography study. There was no difference by latitude in incidence. In the previous study incidence increased towards the north.

**Table 4.3. Univariate Poisson regression analysis of risk factors for reported campylobacteriosis cases.**

| Variable (risk factor) | Unit | Estimate of regression coefficient ($\beta$) | Std. Error | P-value |
|---|---|---|---|---|
| Latitude | degree | -0.014 | 0.0075 | 0.068 |
| Longitude | degree | 0.036 | 0.0064 | <0.001 |
| SIMDScore* | - | -0.006 | 0.0003 | <0.001 |
| Human population density | people/km$^2$ | $-2.14\times10^{-5}$ | $1.44\times10^{-6}$ | <0.001 |
| Private water density | Number of properties / number of people | 0.172 | 0.1481 | 0.246 |
| Poultry density | poultry/km$^2$ | $-2.15\times10^{-5}$ | $3.03\times10^{-6}$ | <0.001 |
| Cattle density | cattle/km$^2$ | 0.001 | 0.0002 | <0.001 |
| Sheep density | sheep/km$^2$ | 0.001 | $8.09\times10^{-5}$ | <0.001 |

If the regression coefficient ($\beta$) is positive the incidence increases as the risk factor increases and if it is negative it decreases (for brevity the intercepts are not provided). The p-values indicate significance. The red colour indicates a significant increase in incidence when the risk factor increases, whilst blue is the opposite showing a significant decrease and black shows no significant difference. *As deprivation increases then campylobacteriosis incidence decreases.

(ii) Multivariate Poisson regression

The multivariate Poisson regression analysis (Table 4.4) shows that increasing deprivation, latitude (i.e. further north), human population density, PWS density, poultry density and cattle density was associated with decreasing incidence (i.e. "protective" for human campylobacteriosis). For cattle density the result is opposite to the finding from the univariate analysis (such a reversal can occur based on the relationship between variables in the model). Both longitude (west to east) and sheep density were associated with increased risk of human campylobacteriosis as in the univariate analysis. In the previous geography study only increasing deprivation (Carstairs index) and increasing human population density were associated with increased risk of human campylobacteriosis.

**Table 4.4. Multivariate Poisson regression analysis of risk factors for reported campylobacteriosis cases.**

| Variable (risk factor) | Unit | Estimate of regression coefficient ($\beta$) | Std. Error | p-value |
|---|---|---|---|---|
| Intercept | | -0.601 | 0.5107 | 0.240 |
| Latitude | degree | -0.071 | 0.0088 | <0.001 |
| Longitude | degree | 0.046 | 0.0073 | <0.001 |
| SIMDScore* | - | -0.005 | 0.0004 | <0.001 |
| Human population density | people/km$^2$ | $-2.07\times10^{-5}$ | $1.59\times10^{-6}$ | <0.001 |
| Private water density | Number of properties / number of people | -0.380 | 0.161 | 0.018 |
| Poultry density | poultry/km$^2$ | $-2.77\times10^{-5}$ | $3.12\times10^{-6}$ | <0.001 |
| Cattle density | cattle/km$^2$ | -0.001 | 0.0002 | 0.002 |
| Sheep density | sheep/km$^2$ | 0.000 | $9.55\times10^{-5}$ | <0.001 |

If the regression coefficient ($\beta$) is positive the incidence increases as the risk factor increases and if it is negative it decreases. The p-values indicate significance. The red colour indicates a significant increase in incidence when the risk factor increases, whilst blue is the opposite showing a significant decrease and black shows no significant difference. *As deprivation increases then campylobacteriosis incidence decreases.

#### 4.3.3.3 Risk factors associated with human campylobacteriosis in Scotland. Results from univariate and multivariate binary logistic regression.

**(i) Univariate logistic regression**

Table 4.5 presents the results from the univariate binary logistic regression comparing least (SIMD5) and most (SIMD1) deprived quintiles for each risk factor.

Note that most of the results presented in this section do not control for denominator populations. Thus, to a first approximation, differences might simply reflect different numbers of people in each category (see Example 3 below). It is thus difficult to interpret these results in isolation, in terms of relevance of the factor considered and to ascribe significance to it in terms of, say, solely campylobacteriosis risk.

**Table 4.5. Univariate binary logistic regression comparing cases in least deprived and most deprived data zones**

| Risk factor | Estimate of regression coefficient ($\beta$) | Std. Error | OR(95% CI) | P-value |
|---|---|---|---|---|
| GENDER | | | | |
| Male (reference) | | | | |
| Female | -0.117 | 0.034 | 0.890(0.832, 0.951) | 0.001 |
| POPULATION DENSITY | | | | |
| Urban - High population density (reference) | | | | |
| peri-Urban - Intermediate population density | 0.752 | 0.0043 | 2.121(1.950, 2.307) | <0.001 |
| Rural - Low population density | 1.800 | 0.100 | 6.049(4.974, 7.355) | <0.001 |
| POSITION (continuous variable) | | | | |
| Longitude | 0.862 | 0.026 | 2.368(2.250, 2.492) | <0.001 |
| Latitude | 0.947 | 0.041 | 2.577(2.379, 2.791) | <0.001 |
| AGE | | | | |
| 65+ years old (reference) | | | | |
| 0-4 years old | -0.781 | 0.088 | 0.458(0.385, 0.544) | <0.001 |
| 5-24 years old | 0.194 | 0.059 | 1.214(1.082, 1.361) | 0.001 |
| 25-64 years old | -0.233 | 0.043 | 0.792(0.728, 0.861) | <0.001 |
| PRIVATE WATER SUPPLY | | | | |
| Properties on PWS (No) (reference) | | | | |
| Properties on PWS (Yes) | 1.861 | 0.103 | 6.431(5.254, 7.870) | <0.001 |
| CATTLE DENSITY | | | | |
| Cattle density-High (reference) | | | | |
| Cattle density-Low | 0.004 | 0.049 | 1.004(0.911, 1.106) | 0.941 |
| Cattle density-Mid1 | -0.544 | 0.049 | 0.580(0.527, 0.639) | <0.001 |
| Cattle density-Mid2 | 0.003 | 0.052 | 1.003(0.907, 1.110) | 0.947 |

| Table 4.5 (contd.) Risk factor | Estimate of regression coefficient ($\beta$) | Std. Error | OR(95% CI) | P-value |
|---|---|---|---|---|
| SHEEP DENSITY | | | | |
| Sheep density-High (reference) | | | | |
| Sheep density-Low | -0.491 | 0.051 | 0.612(0.554, 0.676) | <0.001 |
| Sheep density-Mid1 | -0.382 | 0.053 | 0.683(0.615, 0.757) | <0.001 |
| Sheep density-Mid2 | -0.034 | 0.055 | 0.967(0.868, 1.077) | 0.540 |
| POULTRY DENSITY | | | | |
| Poultry density-High (reference) | | | | |
| Poultry density-Low | -0.107 | 0.049 | 0.899(0.817, 0.989) | 0.028 |
| Poultry density-Mid1 | -0.148 | 0.051 | 0.863(0.781, 0.953) | 0.004 |
| Poultry density-Mid2 | 0.024 | 0.052 | 1.024(0.926, 1.133) | .0640 |
| TIME OF YEAR | | | | |
| Rest of year (reference) | | | | |
| Summer (May, Jun, Jul, Aug) | 0.014 | 0.035 | 1.014(0.948, 1.085) | 0.683 |

| Table 4.5 (contd.) Risk factor | | Estimate of regression coefficient ($\beta$) | Std. Error | OR(95% CI) | P-value |
|---|---|---|---|---|---|
| HEALTH BOARD | | | | | |
| | TY(reference) | | | | |
| | AA | -0.789 | 0.088 | 0.454(0.382, 0.540) | <0.001 |
| | BR | 0.525 | 0.209 | 1.691(1.121, 2.549) | 0.012 |
| | DG | 0.186 | 0.184 | 1.205(0.840, 1.729) | 0.311 |
| | FF | 0.009 | 0.100 | 1.009(0.830, 1.226) | 0.929 |
| | FV | 0.078 | 0.093 | 1.081(0.901, 1.297) | 0.402 |
| | GC | -0.620 | 0.068 | 0.538(0.471, 0.614) | <0.001 |
| | GR | 1.704 | 0.100 | 5.495(4.519, 6.681) | <0.001 |
| | HG | -0.140 | 0.125 | 0.869(0.680, 1.110) | 0.125 |
| | LN | -0.733 | 0.076 | 0.480(0.414, 0.558) | <0.001 |
| | LO | 0.891 | 0.076 | 2.439(2.100, 2.831) | <0.001 |

For brevity the intercepts are not provided. Significant results are coloured in red (comparison of least deprived compared to most deprived is significantly higher than the reference) and blue (comparison is significantly lower).

For ease of interpretation results for three examples are now explained in detail.

Example 1, considers *gender* as a risk factor. The ratio of the number of female cases residing in least deprived areas to the number in most deprived areas, is compared with the same (reference) ratio for males. The odds ratio and corresponding P- values are determined by univariate logistic regression (Table 4.5). Since the OR <1 (i.e. =0.890) this means that the number of female cases in least deprived areas is proportionally lower than for male (*reference*) cases.

Example 2, considers *human population density* as a risk factor (where urban population is the reference). Since the peri-urban population has an OR>1 this means that the number of cases in least deprived peri-urban areas is proportionally higher than that for urban (*reference*) areas. The same occurs for the rural comparison.

Example 3, investigates *longitude* as a risk factor. In the above two examples the risk factors were treated as categorical variables. Here longitude is described as a continuous variable. Since the OR>1 (i.e. =2.368) this means that the ratio of least deprived cases/most deprived cases, increases by a factor of 2.368 towards the east in Scotland for each degree increase in longitude (a degree corresponds to approximately 50 miles in Scotland). Hence, there are proportionally more cases in least deprived East Scotland data zones compared with West Scotland. This is, perhaps, unsurprising as there are more *people* living in least deprived areas, in the East of Scotland.

The remaining risk factors in Table 4.5 are discussed below.

*Latitude*: the ratio of the number of cases in least deprived to number in most deprived areas, in the north is higher than in the south (i.e. there are proportionally more cases in least deprived northerly areas).

*Age*: for both 0-4 and 25-64 year-old cases the ratio of the number in least deprived to number in most deprived areas, is significantly lower than for 65+ years old cases (the reference group). This means that in 0-4 and 25-64 year-old cases, there are proportionally more in the most deprived areas than for 65+ year-olds. The situation is opposite for the 5-24 year old age group, where the number of cases in least deprived areas is proportionally higher than for the 65+ year-olds.

*PWS*: the ratio of the number of cases in least deprived to number in most deprived data zones served by PWS, is higher than that for data zones not served by PWS. This means that there are proportionally more cases in least deprived data zones where there are properties on PWS. Again, this may not be surprising if there are more *people* in least deprived data zones where there are properties on PWS.

*Cattle density*: the ratio of the number of cases in least deprived to number in most deprived data zones with "Mid1" (intermediate) cattle density is lower than for areas with high cattle density. This means that the number of cases in least deprived data zones with "Mid1" cattle density is proportionally lower than for areas with high cattle density. There were no other significant differences. Hence, the results *appear* to be inconsistent as it would be expected that "Low" cattle density would also be significant. (Note however, that calculating

population rates might clarify this finding depending how *people* are distributed between these data zone categories.)

*Sheep density*: there is a decreasing trend in the ratio of the number of cases in least deprived to number in most deprived areas, as sheep density decreases. So in areas of higher sheep densities there are proportionally more cases in least deprived areas. It is unclear why this should be the case.

*Poultry density*: the ratio of the number of cases in least deprived to number in most deprived data zones, with "Low" and "Mid1" poultry densities is lower than for data zones with high poultry density (the reference group). So there are proportionately more cases in least deprived areas of higher poultry density. This is the same for sheep.

*Time of year*: the ratio of the number of cases from least deprived to number from most deprived areas did not vary during the year.

*Health board*: Tayside health board had the highest incidence (143.6 cases/100,000 (95%CI – 129.5 – 157.7)) during the time period between 1st January 2012 to 31st March 2018 and was used as the reference in the logistic regression analysis when comparing health boards.

The ratios of the number of cases from least deprived to number from most deprived areas of AA, GC and LN are lower than for Tayside. This means that in these health boards there are proportionally fewer cases in least deprived data zones than in Tayside. The situation is opposite for BR, GR and LO, where there are proportionally more least deprived cases than in Tayside. These differences may be due to a different distribution of the human population in each health board. For example, Grampian will have a greater proportion of the population living in least deprived data zones. Hence, it is likely there will be more cases in this group simply because of the larger population in least deprived areas.

**(ii) Multivariate logistic regression**

All factors having a P-value <0.25 in the univariate analysis were introduced into the multivariate analysis simultaneously. Table 4.6 provides the results and the method of interpretation is similar to the univariate analysis.

**Table 4.6. Multivariate binary logistic regression comparing cases in the least deprived and most deprived data zones.**

| Risk factor | Estimate of regression coefficient ($\beta$) | Std. Error | OR(95% CIs) | P-value |
|---|---|---|---|---|
| GENDER | | | | |
| Male (reference) | | | | |
| Female | -0.110 | 0.039 | 0.896(0.831, 0.966) | 0.004 |
| POPULATION DENSITY | | | | |
| Urban - High population density (reference) | | | | |
| peri-Urban - Intermediate population density | 0.967 | 0.125 | 2.631(2.058, 3.363) | <0.001 |
| Rural - Low population density | 0.873 | 0.049 | 2.394(2.176, 2.633) | <0.001 |
| POSITION (continuous variable) | | | | |
| Longitude | -0.032 | 0.103 | 0.969(0.791, 1.186) | 0.760 |
| Latitude | -0.507 | 0.130 | 0.602(0.467, 0.776) | <0.001 |
| AGE | | | | |
| 65+ years old (reference) | | | | |
| 0-4 years old | -1.020 | 0.102 | 0.361(0.295, 0.440) | <0.001 |
| 5-24 years old | 0.035 | 0.066 | 1.036(0.910, 1.178) | 0.594 |
| 25-64 years old | -0.262 | 0.048 | 0.770(0.701, 0.845) | <0.001 |
| PRIVATE WATER SUPPLY | | | | |
| Properties on PWS (No) (reference) | | | | |
| Properties on PWS (Yes) | 0.661 | 0.133 | 1.937(1.492, 2.516) | <0.001 |
| CATTLE DENSITY | | | | |
| Cattle density-High (reference) | | | | |
| Cattle density-Low | -0.125 | 0.102 | 0.882(0.722, 1.078) | 0.220 |
| Cattle density-Mid1 | -0.700 | 0.078 | 0.496(0.426, 0.579) | <0.001 |
| Cattle density-Mid2 | -0.246 | 0.070 | 0.782(0.682, 0.898) | <0.001 |

| (continued) Risk factor | | Estimate of regression coefficient ($\beta$) | Std. Error | OR(95% CIs) | P-value |
|---|---|---|---|---|---|
| SHEEP DENSITY | | | | | |
| | Sheep density-High (reference) | | | | |
| | Sheep density-Low | -1.257 | 0.096 | 0.284(0.236, 0.343) | <0.001 |
| | Sheep density-Mid1 | -0.410 | 0.071 | 0.663(0.577, 0.763) | <0.001 |
| | Sheep density-Mid2 | 0.072 | 0.066 | 1.075(.944, 1.224) | 0.275 |
| POULTRY DENSITY | | | | | |
| | Poultry density-High (reference) | | | | |
| | Poultry density-Low | 1.183 | 0.082 | 3.263(2.778, 3.833) | <0.001 |
| | Poultry density-Mid1 | 0.470 | 0.072 | 1.599(1.389, 1.841) | <0.001 |
| | Poultry density-Mid2 | 0.991 | 0.074 | 2.695(2.332, 3.114) | <0.001 |
| HEALTH BOARD | TY(reference) | | | | |
| | AA | -2.513 | 0.202 | 0.081(.055, .120) | <0.001 |
| | BR | -0.688 | 0.251 | 0.503(.307, .822) | .006 |
| | DG | -1.592 | 0.263 | 0.204(.122, .341) | <0.001 |
| | FF | -0.425 | 0.117 | 0.654(.519, .823) | <0.001 |
| | FV | -0.839 | 0.130 | 0.432(.335, .558) | <0.001 |
| | GC | -1.581 | 0.162 | 0.206(.150, .283) | <0.001 |
| | GR | 1.628 | 0.154 | 5.092(3.764, 6.887) | <0.001 |

| | | | | |
|---|---|---|---|---|
| HG | -0.930 | 0.220 | 0.394(.256, .607) | <0.001 |
| LN | -1.964 | 0.147 | 0.140(.105, .187) | <0.001 |
| LO | 0.265 | 0.111 | 1.303(1.048, 1.620) | .017 |
| Intercept** | 29.538 | 7.405 | 6.7E+12(na*, na) | <0.001 |

References are as in univariate and indicated in the table. Significant results are coloured in red (comparison is significantly higher than the reference) and blue (comparison is significantly lower).

* na – not applicable

**The intercept in the logistic regression sets the "baseline" event rate, i.e. the natural logarithm of the odds ratio when all risk factors values are set equal to zero simultaneously (http://www.med.mcgill.ca/epidemiology/joseph/courses/EPIB-621/logistic2.pdf ). In practice when there are more than two risk factors (covariates) it is unlikely to have them all set at zero simultaneously. Hence in the above multivariate logistic regression the intercept has no physical meaning. However, using an intercept in the logistic regression is important, otherwise the model will be forced through the origin.

The main findings from Table 4.6 are:

*Gender, human population density, properties on PWS* and *sheep density*: The results from the multivariate analysis were either the same or very similar to the univariate analysis. *Age* gave similar results while *cattle* and *longitude results* were inconsistent.

*Health board*: The ratio of the number of least deprived cases/number of most deprived cases, in AA, BR, DG, FF, FV, GC, HG and LN is significantly lower than in Tayside (the reference health board). This means that in these health boards there are proportionally fewer least deprived cases than in Tayside. The situation is opposite for GR and LO. This is largely unsurprising as there are fewer less deprived people/data zones in these NHS board areas compared with the reference (TY). The health boards that were not significant in the univariate analysis have now become significant with proportionally fewer least deprived cases than Tayside.

*Position*: Longitude (west to east) is no longer significant. But latitude has now reversed - there are proportionally fewer least deprived cases as you go north.

*Poultry density*: This is opposite to the finding from the univariate analysis. So, after accounting for other factors, there are proportionately fewer cases in least deprived areas of higher poultry density.

#### 4.3.3.4 Risk factors associated with human campylobacteriosis in Scotland. Results from univariate and multivariate multinomial logistic regression

**(i) Results from multinomial univariate logistic regression between cases classified by SIMD quintile**

The results presented in Figure 4.11 compare the univariate multinomial logistic regression from *less* deprived (SIMD5, 4, 3 & 2) and *most* (SIMD1) deprived quintiles for each risk factor. The results can be challenging to interpret and hence for ease of interpretation results for two examples are now explained in detail.

Example 1 investigates *gender* as a risk factor, where male is considered as the reference (see the "r" letter above "Male" bars in Figure 4.11 (a)). Based on eq. (4.10) an odds ratio was calculated to compare female vs. male cases between SIMD5 & 1:

$$OR_{female,SIMD5v1} = \frac{\left(\frac{SIMD5\ cases}{SIMD1\ cases}\right)_{female}}{\left(\frac{SIMD5\ cases}{SIMD1\ cases}\right)_{male}} \qquad (4.10)$$

This represents the ratio of the number of SIMD5 cases/number of SIMD1 cases, in females divided by the number of SIMD5 cases/number of SIMD1 cases, in males. This odds ratio is presented in Figure 4.11(a) as a purple bar for female. The counterpart purple bar for male equals 1 (this is the reference where males are compared with males). Since $OR_{female,SIMD5v1}$ <1 (i.e. purple bar for female is below the red line in Figure 4.11(a)) this means that the number of SIMD5 cases

in female is proportionally lower than that in the reference male population. The "-" blue sign above the purple female bar indicates that this is statistically significant.

Using equations (4.6) to (4.8), odds ratios comparing female with male were calculated for cases in SIMD4, 3 & 2 quintiles compared with SIMD1. These odds ratios are represented by the green, red and blue bars (Figure 4.11(a)). There are proportionally fewer SIMD4 and SIMD3 cases in female than in male (Odds ratio's <1), whilst for SIMD2 there is no significant difference between female and male (note that the blue bar is close to 1).

Example 2 considers the *human population density* as a risk factor (urban population as reference). An odds ratio as described in equation (4.10) can be calculated to compare rural vs. urban cases between SIMD 5 & 1:

$$OR_{rural,SIMD5v1} = \frac{\left(\frac{SIMD5\ cases}{SIMD1\ cases}\right)_{rural}}{\left(\frac{SIMD5\ cases}{SIMD1\ cases}\right)_{urban}} \qquad (4.11)$$

Now the $OR_{rural,SIMD5v1}$>1 is given by the purple bar for "rural" in Figure 4.11(b). This means that the number of SIMD5 cases in the rural population is proportionally higher than that in the urban reference population. The "+" red sign above the purple "Rural" bar indicates this is statistically significant. Similar results occur for the SIMD4, 3 & 2 comparisons. Hence, there are proportionally more cases in the *less* deprived (SIMD5, 4, 3 & 2) rural areas than in the urban ones. The same occurs for peri-Urban areas.

The remaining risk factors in Figure 4.11 are discussed below.

*Longitude:* There are proportionally more cases in the *less* deprived East of Scotland areas than in those of the West of Scotland (Figure 4.11(c)).

*Latitude:* There are proportionally more cases in *less* deprived areas of the North than in the South of Scotland (Figure 4.11(d)).

*Age:* There are proportionally fewer 0-4 and 25-64 years old cases in *less* deprived (SIMD5, 4, 3 & 2) areas compared to 65+ years old (Figure 4.11(e)). (For the 5-24 years old age group the results are inconsistent.)

*PWS:* There are proportionally fewer cases in *less* deprived (SIMD5, 4, 3 & 2) data zones where PWS are not present than for data zones with PWS (Figure 4.11(f)).

*Cattle density:* There are proportionally fewer cases in *less* deprived (SIMD5, 4, 3 & 2) data zones with intermediate "Mid1" cattle density than in those with high cattle density (Figure 4.11(g)). This was true also for data zones with "Low" cattle density except for SIMD5.

*Sheep density:* There are proportionally fewer cases in *less* deprived (SIMD5, 4, 3 & 2) data zones with "Low" and intermediate ("Mid1" & "Mid2") sheep densities than in data zones with high sheep density (Figure 4.11(h)). So there are proportionally more cases in *less* deprived parts of high sheep density regions.

*Poultry density:* There are proportionally fewer cases in *less* deprived (SIMD5, 4, 3 & 2) data zones with low and intermediate ("Mid1") poultry density than in

data zones with high poultry density (Figure 4.11(i)). The results are also true for data zones with "Mid2" poultry density except for SIMD5.

*Time of year (season):* There are proportionally fewer SIMD3 cases during the rest of year than during the "summer" months (Figure 4.11(j)). However, there is no difference for the other comparisons.

*Health board:* As for binomial logistic regression, Tayside health board was used as the reference in the multinomial logistic regression analysis (Figure 4.11(k)).

The results are presented below for each health board.

AA: There are proportionally fewer (SIMD5, 4 & 3) cases in AA than in TY. There is no difference for the SIMD2 cases.

BR: There are proportionally more cases in *less* deprived (SIMD5, 4, 3 & 2) areas of BR than for TY.

DG: There are proportionally more (SIMD4, 3 & 2) cases in DG than in TY. There is no difference for the SIMD5 cases.

FF: There are proportionally fewer (SIMD4 & 3) cases in FF than in TY. There is no difference for the SIMD5 & 2 cases.

FV: There are proportional fewer SIMD4 cases in FV than in TY. Also, there are proportional more SIMD2 cases in FV than in TY. There is no difference for the SIMD5 & 3 cases. Hence the results for FV are inconclusive.

GC: There are proportionally fewer cases in *less* deprived (SIMD5, 4, 3 & 2) areas of GC than for Tayside.

GR: There are proportionally more cases in *less* deprived (SIMD5, 4, 3 & 2) areas of GR than for TY.

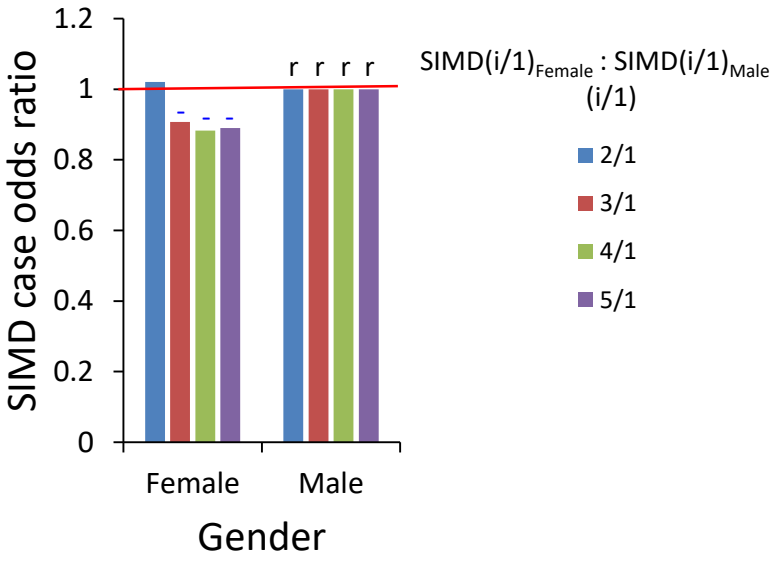
HG: There are proportionally more cases in (SIMD4, 3 & 2) areas of HG than for TY. There is no difference for the SIMD5 cases.

LN: There are proportionally fewer cases in (SIMD5, 4 & 3) areas of LN than for TY. The result is opposite for the SIMD2 cases. Hence, the results for LN are inconclusive.

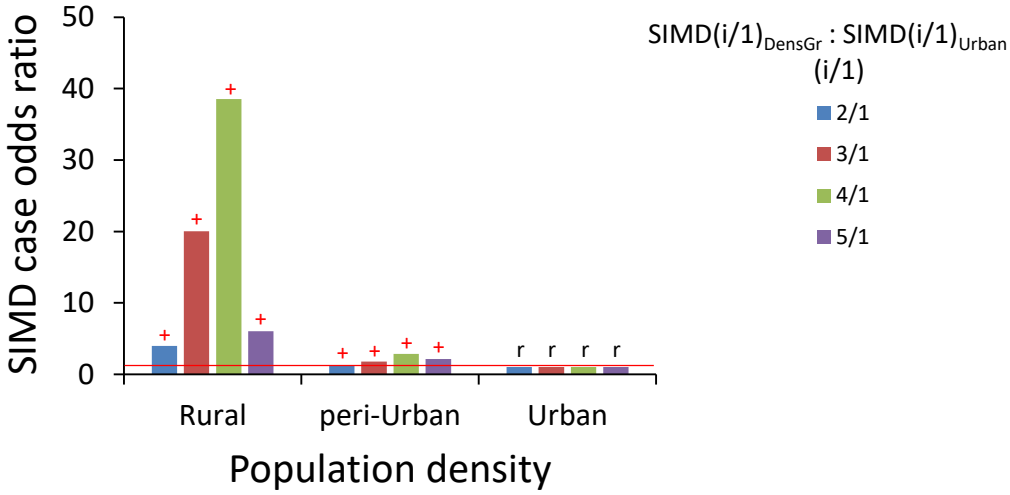
LO: There are proportionally more cases in SIMD5, 3 & 2 areas of LO than for TY. This is not true for the SIMD4 cases. Hence, the results for LO are inconclusive.

**Figure 4.11. Multinomial univariate logistic regression comparing cases in less deprived quintiles with most deprived quintiles, for each risk factor.**

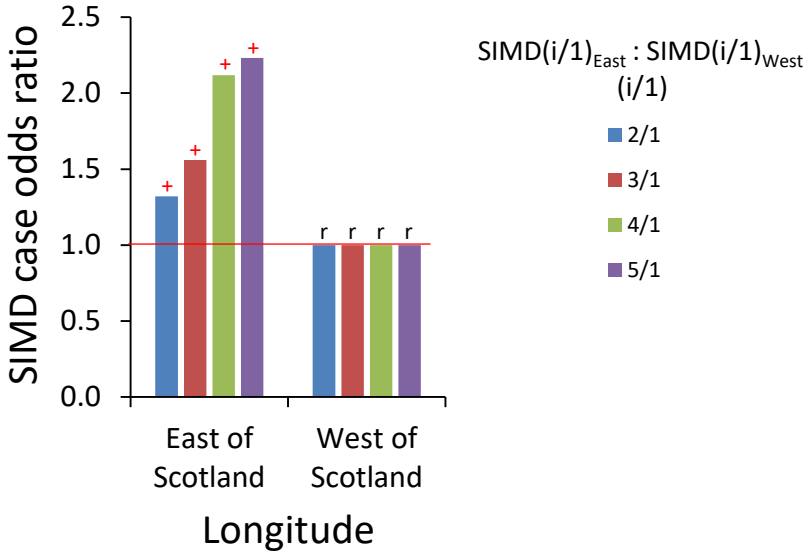
(a)



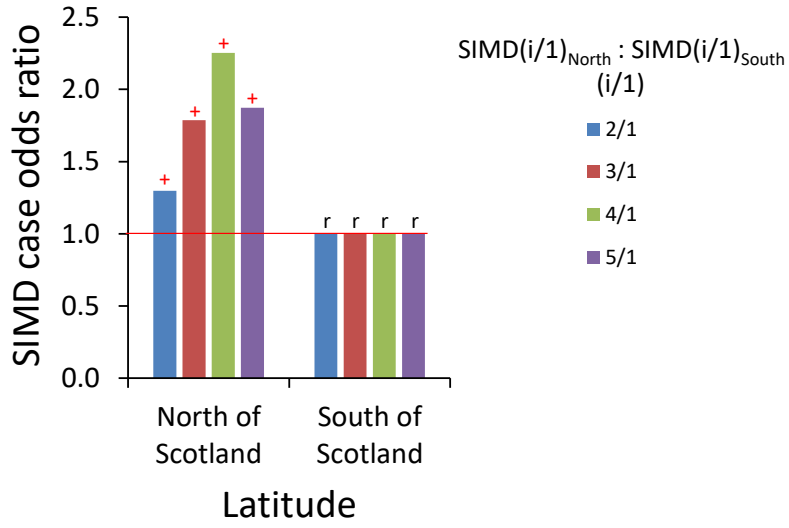

(b)



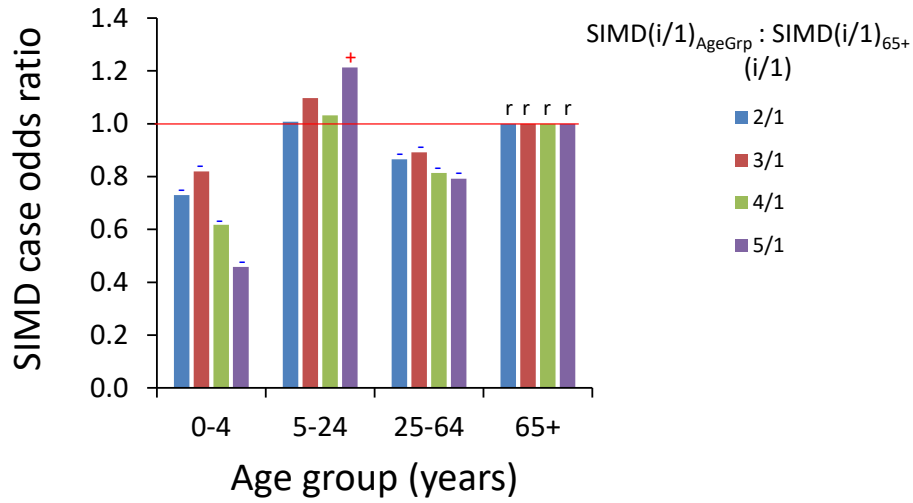

(c)



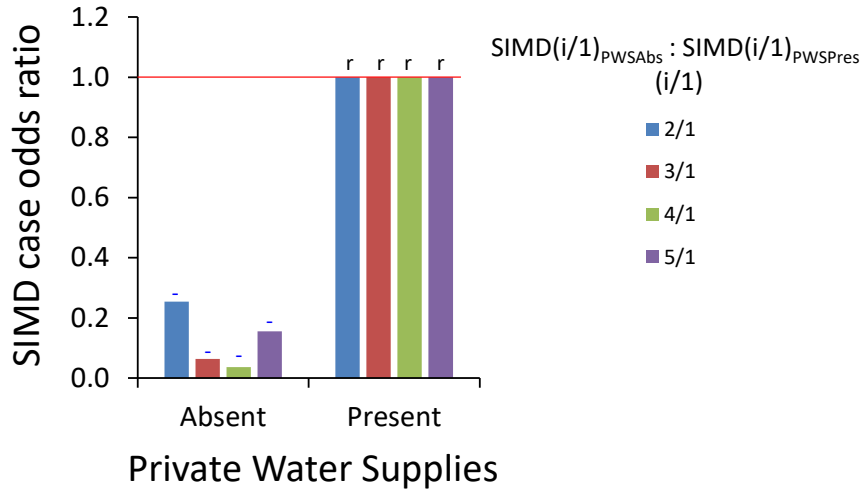
(d)


SIMD case odds ratio
2.5
2.0
1.5
1.0
0.5
0.0
North of Scotland
South of Scotland
Latitude
SIMD(i/1)North : SIMD(i/1)South (i/1)
2/1
3/1
4/1
5/1

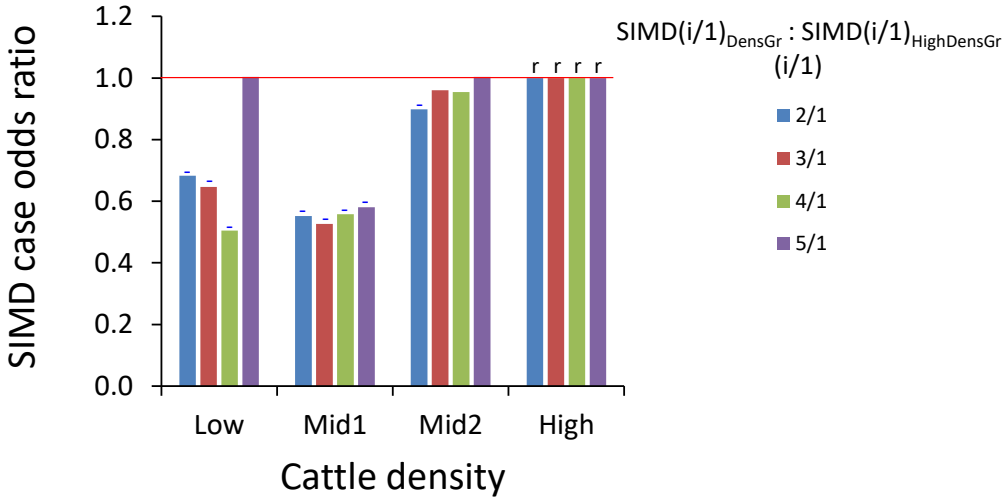
(e)


SIMD case odds ratio
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0-4
5-24
25-64
65+
Age group (years)
SIMD(i/1)AgeGrp : SIMD(i/1)65+ (i/1)
2/1
3/1
4/1
5/1

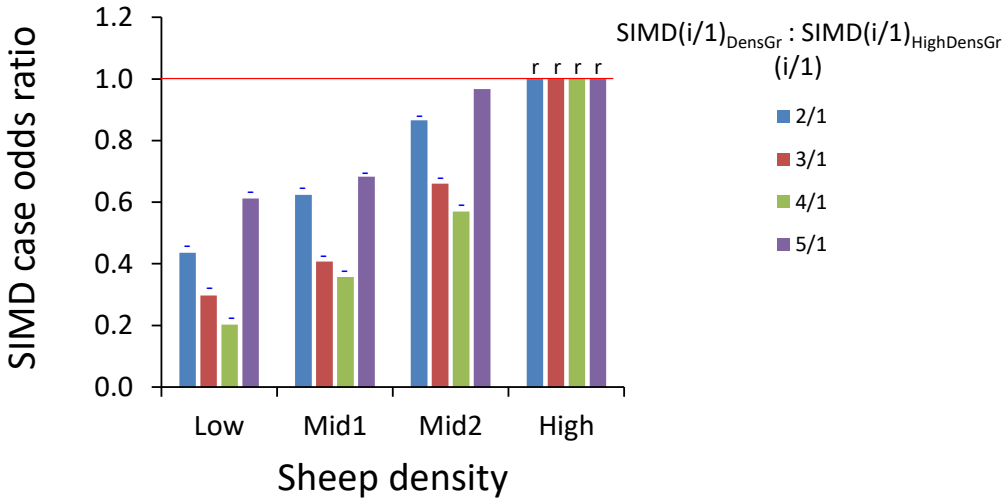
(f)


SIMD case odds ratio
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Absent
Present
Private Water Supplies
SIMD(i/1)PWSAbs : SIMD(i/1)PWSPres (i/1)
2/1
3/1
4/1
5/1

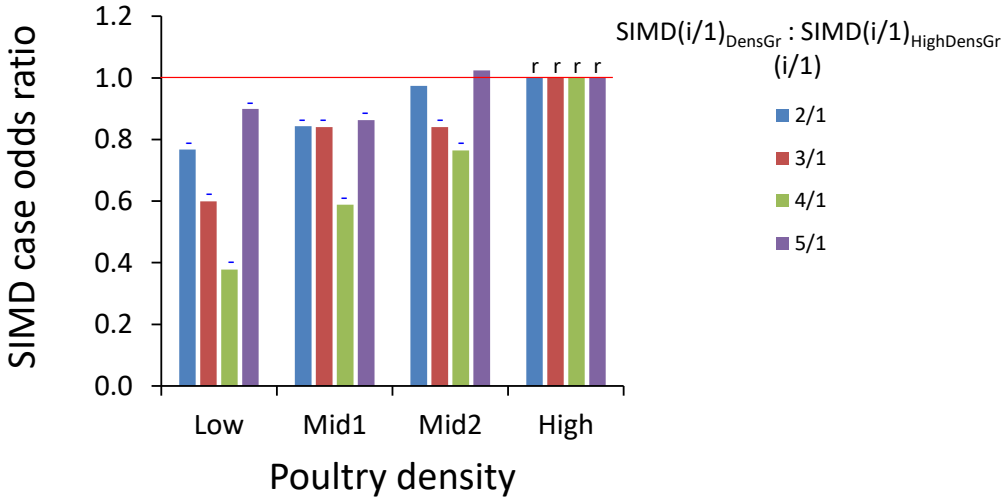
(g)


SIMD case odds ratio
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Low
Mid1
Mid2
High
Cattle density
SIMD(i/1)DensGr : SIMD(i/1)HighDensGr (i/1)
2/1
3/1
4/1
5/1

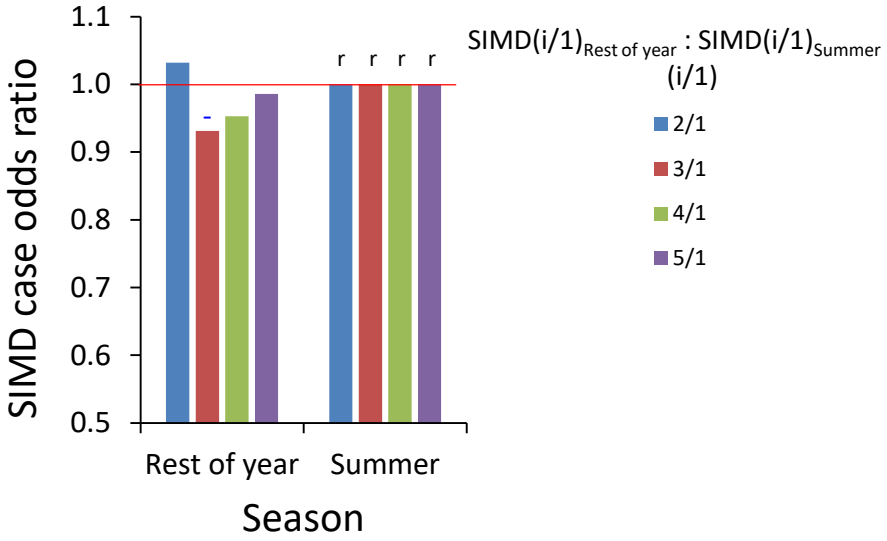
(h)


SIMD case odds ratio
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Low
Mid1
Mid2
High
Sheep density
SIMD(i/1)DensGr : SIMD(i/1)HighDensGr (i/1)
2/1
3/1
4/1
5/1

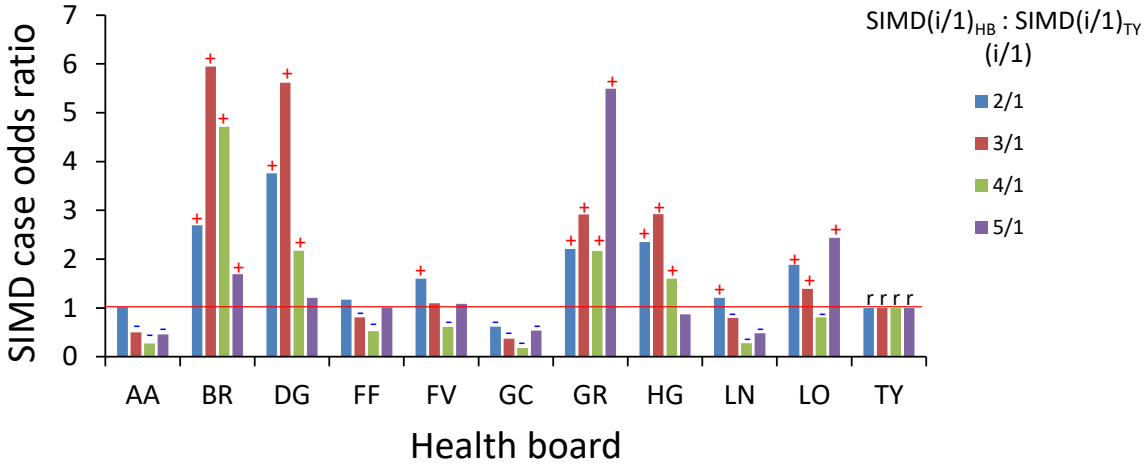
(i)


SIMD case odds ratio
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Low
Mid1
Mid2
High
Poultry density
SIMD(i/1)DensGr : SIMD(i/1)HighDensGr (i/1)
2/1
3/1
4/1
5/1

(j)




(k)




Multinomial univariate logistic regression comparing cases in *less* deprived (SIMD5, 4, 3 & 2) quintiles with most deprived (SIMD1) quintiles, for each risk factor. (a) gender, (b) population density, (c) longitude, (d) latitude, (e) age, (f) PWS, (g) cattle density, (h) sheep density, (i) poultry density, (j) time of year (season) and (k) health board. Where SMID5 is least deprived and SIMD1 is most deprived. The letter "r" denotes the reference and "+" indicates the comparison is significantly higher whilst "-" indicates that it is significantly lower.

**(ii) Results from multinomial multivariate logistic regression between cases classified by SIMD quintile**

These analyses are presented in Annex 4.2.

The risk factor *time of year* was removed during the analysis because it was not significant.

The following risk factors gave the same results as in the univariate analysis for all SIMD comparisons: *human population density* and *health board* (GC vs. TY only).

For all the other risk factors, there were some differences from the univariate analysis and these are provided in Annex 4.2.

## 4.4 Perform analysis on spatial distribution of reported campylobacteriosis cases relative to the position of GP practices

### 4.4.1 Aims

The aim of this section was to answer the following questions:

(i) "Are you more likely to be reported for campylobacteriosis if you live close to a GP practice?".
(ii) Does this depend on deprivation?

### 4.4.2 Data

Reported case data from ECOSS were available for the period 1st January 2012 to 31st March 2018 as discussed in Chapter 4 sections 4.2 & 4.3. The human population in each data zone was obtained from The Consumer Data Research Centre (https://data.cdrc.ac.uk/dataset/cdrc-2011-population-weighted-centroids-gb). This also provided coordinates (easting and northing) of the centroid of each data zone. The SIMD quintiles for each data zone were obtained from http://www.gov.scot/Topics/Statistics/SIMD/DataAnalysis/Background-Data-2012.

The easting and northing as well as deprivation quintile of each case and each individual person of the Scottish population was then allocated.

The postcodes of each GP practice (n=950, 2017) in Scotland was obtained from ISD (http://www.isdscotland.org/Health-Topics/General-Practice/Workforce-and-Practice-Populations/) and was geocoded (easting and northing) using the UK Grid Reference Finder (https://gridreferencefinder.com/)

### 4.4.3 Methods

The distance between each case and its closest GP practice was determined. This was plotted as a frequency distribution.

The "control" population comprised the same number of individuals as cases, but was randomly selected from the whole Scottish population. The distance between each "control" and the closest GP practice was then determined. The resulting frequency distribution was compared with that obtained for cases.

This "control" distribution was recalculated 500 times using the Monte Carlo method in PopTools (http://www.poptools.org/). From this average distances from GP practices and 95% confidence intervals were determined. If the confidence intervals did not overlap the case distribution then the results were considered to be significantly different.

This was repeated by SIMD quintile.

### 4.4.4 Results and Discussion

Figure 4.12 shows a map of the 950 GP practices in Scotland. Figure 4.13a shows that the distribution of campylobacteriosis cases around GP practices generally follows the distribution of the "control" population. However there are some significant differences (e.g. at 1km there are fewer campylobacteriosis cases than expected whilst at 2, 6 and 10 km there are more) but these differences are each <1% of reported cases and so are unlikely to have a large impact on overall reporting bias.

For both SIMD1 and SIMD5 the distribution of the reported cases broadly follows the "control" population (Figure 4.13b and c). These graphs illustrate (see Figure 4.1.3b) fewer cases than expected in SIMD1 (most deprived) areas at a GP practice separation distance of 1km). Proximity to a GP practice does not appear to influence the proportion of reported campylobacteriosis cases detected for people living in SIMD5 areas. Also, a higher proportion of the SIMD1 population lives close to a GP practice than SIMD5 (least deprived) population. Thus 1km proximity to a GP practice appears to reduce the proportion of people in SIMD1 (most deprived) data zones who report a case of campylobacteriosis. The difference is small (2%) and insufficient to explain the overall differences in cases reported between SIMD1 and SIMD5 (see Figure 4.2). It is worth noting that the reported case data was at data zone level. Ideally full postcode would be a more precise way of doing this analysis, but this resolution was not available to the authors.

**Figure 4.12. The geographical distribution of GP practices in Scotland (April 2018).**

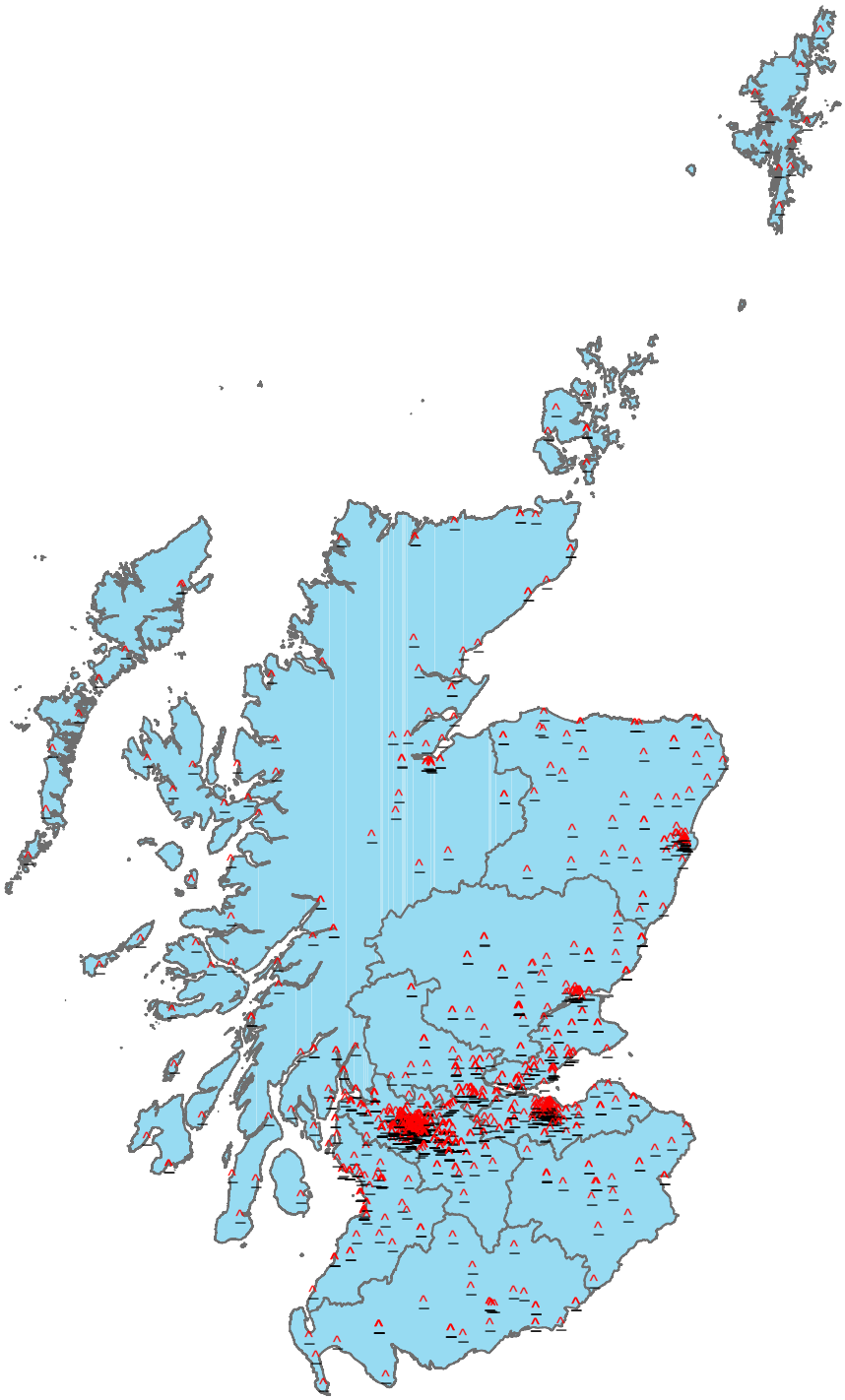
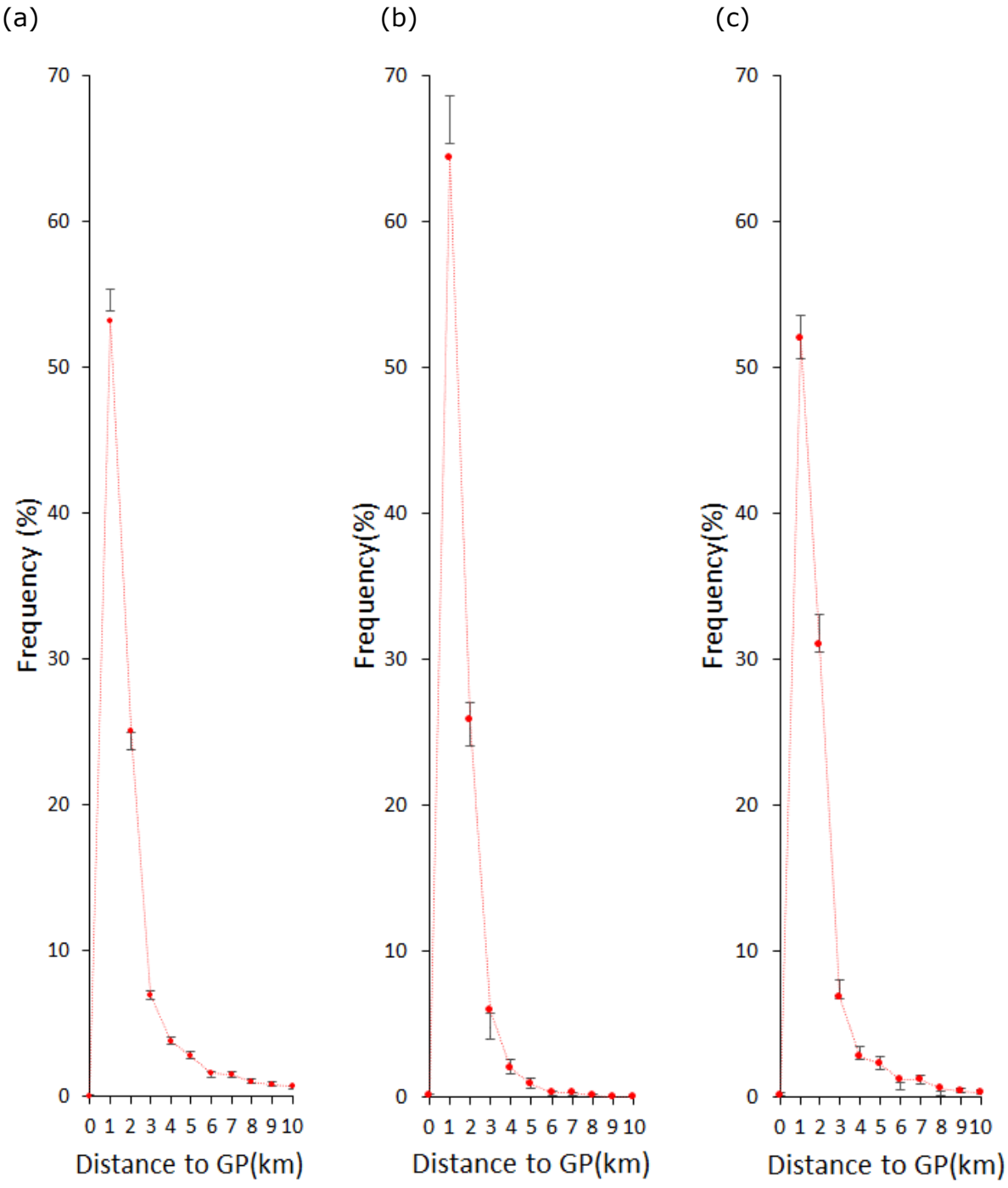


**Figure 4.13. Frequency distribution of campylobacteriosis cases and "control population" relative to the distance to the closest GP practice**




(a) all cases vs population; (b) SIMD1 cases vs. SIMD1 population and (c) SIMD5 cases vs. SIMD5 population. Cases are represented as (●) and confidence intervals are 95 percentiles for the control population (Note: for clarity in the graphs the average distribution of the "control population" is not represented by a symbol).

## 4.5 Long term trends of reported cases in Scotland

### 4.5.1 Aims

The aim of this section is to explore and understand the dynamics of human campylobacteriosis reporting in Scotland over time using time series analysis from long term disease data (1990 to 2017).

### 4.5.2 Data

Summary human campylobacteriosis case data from Scotland during 1990 to 2011 (n=112,230) were available from the literature (N. J. C. Strachan, Rotariu et al. 2013). These data included information about the number of cases in each health board stratified by five year age groups, the incidence (cases/100,000/year) and the ratio of Urban/Rural incidence. Human campylobacteriosis case data from 2012 to 2017 (n=37,611) were available from this study (see sections 4.2 & 4.3) and included information on age and geographical region (i.e. data zone and health board).

Mid-year human population estimates (2012 to 2017) by five year age group, at health board level, were obtained from the National Records of Scotland (NRS) (https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates ). In addition population data, stratified by age were obtained for each data zone from SIMD2012 (http://www.gov.scot/Topics/Statistics/SIMD/DataAnalysis/Background-Data-2012 ).

### 4.5.3 Methods

Campylobacteriosis case data between 1990 to 2017 was plotted in terms of (i) incidence, (ii) incidence stratified by age group and (iii) the Urban/Rural incidence ratio.

To determine the Urban/Rural incidence ratio urban cases were defined as those data zones with >200 persons/km$^2$; other cases were defined as rural. This was done at both data zone and health board geographical resolution. The results from both 1990-2011 and 2012-2017 time periods were combined.

4.5.4 Results and Discussion

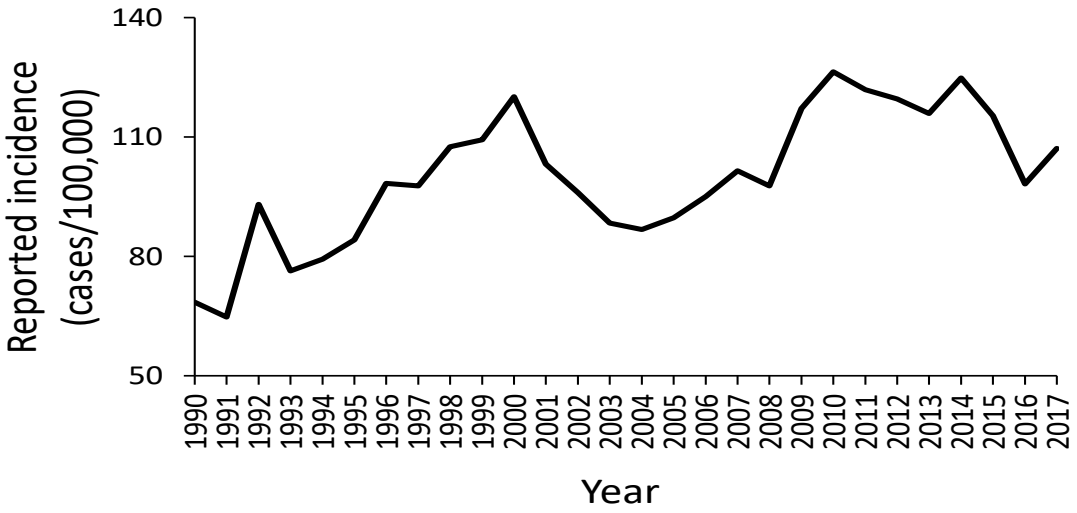
There was an initial peak in the incidence of reported human campylobacteriosis during 2000-01 (Figure 4.14a). This was followed by a decline through 2001-2006. A second peak occurred in 2010 followed by a fall during 2014 and 2016. There has been an increase during 2016 – 2017 which seems surprising because of the efforts to reduce *Campylobacter* in poultry (www.food.gov.uk/news-alerts/news/campylobacter-levels-remain-steady-0 ). However, it has been claimed that due to the ageing population that campylobacteriosis cases would increase because of this (5% between 2010 and 2020) (N. J. C. Strachan, Rotariu et al. 2013).

Reported campylobacteriosis incidence in 0-4 year olds has been decreasing since 1990 (Figure 4.14b). The incidence in 65+ years old increased between 1990 to 2010, after which it has been relatively stable (Figure 4.14(b)). The reasons for the increase are unknown but it has been hypothesised that this may be associated with PPI intake in this population as well as increased consumption of chicken (N. J. C. Strachan, Rotariu et al. 2013).

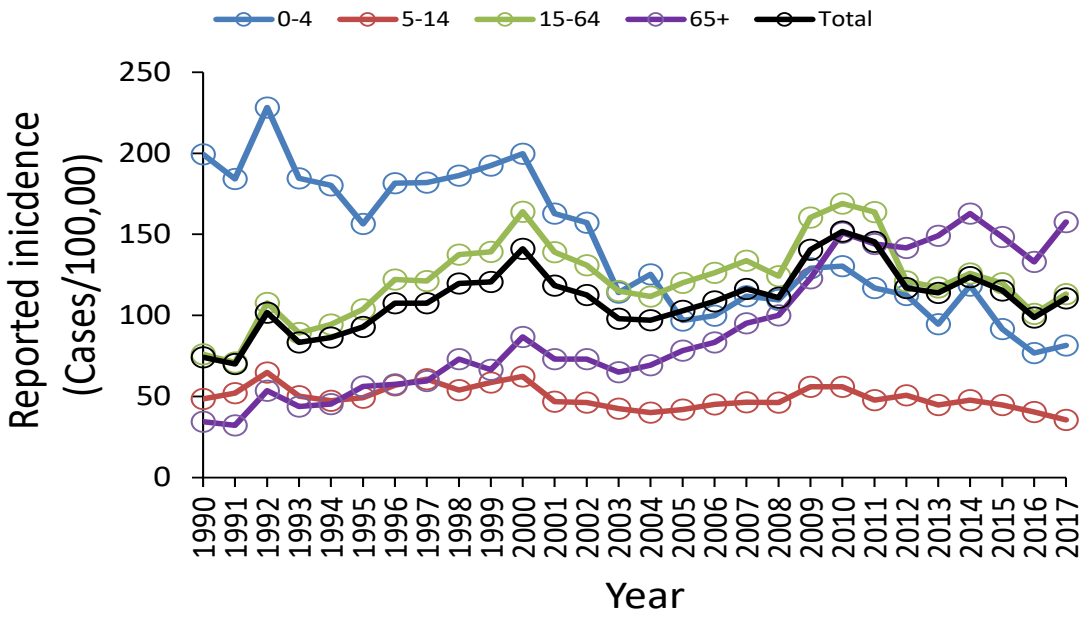
The Urban/Rural (U/R) incidence ratio of reported cases at health board level (Figure 4.14c) has fluctuated during the period under study. It is now higher than it was in the early 1990's. This could be hypothesised to be due to a lower proportion of environmental cases. Since the U/R ratio was calculated from data zones (2012-2017) it is not possible to compare directly with data generated previously as this used postal sectors.

**Figure 4.14. Long term analysis of reported cases**

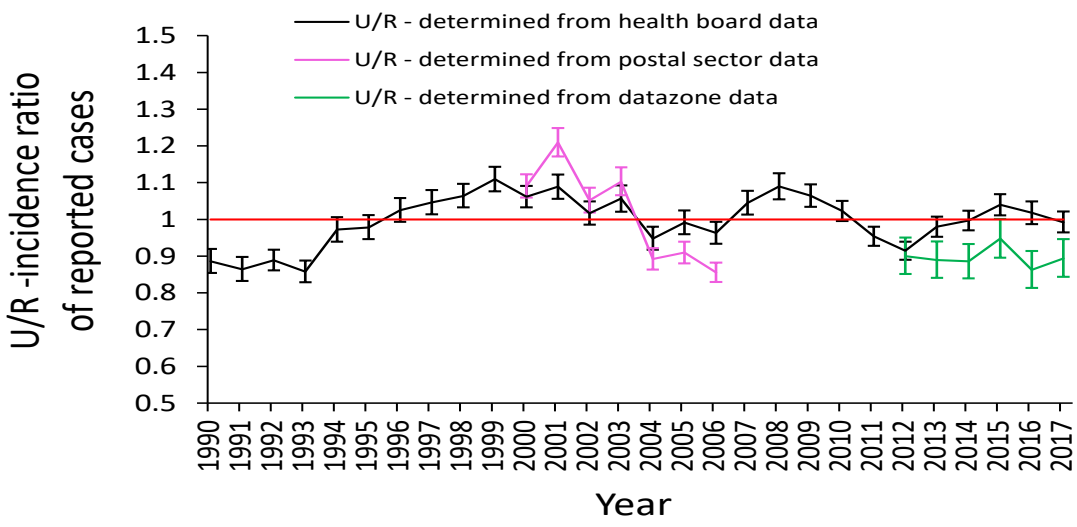
(a)




(b)




(c)




(a) Incidence of reported cases, (b) incidence of reported cases stratified by age and (c) urban/rural incidence ratio of reported cases.

## 4.6 Overall discussion

### 4.6.1 Scottish Population

The Scottish population is concentrated in the central belt and along the east coast (Annex 7.1 Figure 4.1.4). The highest concentrations of deprived data zones are in urban areas, for example in parts of the cities, in particular Glasgow, and in towns across the country (Chapter 1, Figure 1.2). There is also a higher proportion of people in Scotland in the older age groups (>39 years) that are in the least deprived SIMD5 quintile compared with the most deprived SIMD1 quintile (Figure A4.1.6).

### 4.6.2 Descriptive epidemiology

There were more cases (19%) in the four less deprived SIMD quintiles compared with the most deprived SIMD quintile: a 7% reduction from that found during 2000-2006 (Anon. 2007). This excess of cases in the less deprived population has been found in New Zealand (Spencer, Marshall et al. 2012) and England and Wales (Gillespie, O'Brien et al. 2008). It is unclear why this is the case, though it is plausible that some of this may be due to foreign travel and also potentially the type of food that is eaten and the way it is prepared.

Incidence is 11.5% higher in rural compared with urban areas. This was also found in the 2000-06 study. However there are more cases in urban (approx. 1810 per year) compared with rural areas (1433 per year) because more people live in urban areas. More people live in deprived urban areas and this may explain part of the difference between the incidence in urban and rural populations. It has been previously shown that incidence in young children in rural areas is higher (e.g. North-East Scotland (N. J. C. Strachan, Gormley et al. 2009)) and in areas with broiler and dairy operations in the USA (Rosenberg Goldstein, Cruz-Cano et al. 2016). This excess in cases has been thought to be due to environmental exposure (contact with farm/wild animals and their faeces) and/or consumption of water from private supplies (Anon 2010).

Across all ages there tends to be a higher number of cases in the least deprived compared with most deprived quintiles except for young children. It appears that the largest excess is in 0 and 1 year olds (Figure 4.3). This is at a time when the children will be weaned, crawling and sampling their environment whilst teething etc. It is unclear why there is this difference for young children. Most young children living in the most deprived quintile live in urban areas (83%). That being the case, it is less likely that this excess is due to environmental exposure from contact with farm or wild animals. There are at least four other hypotheses for this difference. The first is due to poorer food preparation skills in the home. If that was the case it would be expected that adults would also have a higher rate of disease and this is not observed. The second is that these children are more susceptible to infection because of a lower immune status due to poorer living conditions and diet. The third is that they have a different diet which is more likely to have *Campylobacter* present. And finally, there could be different (denominator) population pyramids for people living in SIMD1 and SIMD5 areas.

At present it is unknown which of these hypotheses may be most relevant and further studies would be required to investigate this.

Health boards do differ in campylobacteriosis incidence. FF has the lowest whilst TY and BR the highest. FF was one of the lower incidence health boards in the previous geography study (2000-06 (Anon. 2007)) and has continued to fall throughout the current study. TY and BR were in the mid-range group of campylobacteriosis incidence from the previous study (approx. 100 cases/100,000) and both have increased to approximately 140 cases/100,000 in the current study. It is unclear whether these changes reflect true incidence rather than sampling, testing and reporting artefacts.

Campylobacteriosis incidence is highest in people older than 50. This is more pronounced than in the previous geography study (Anon. 2007). This follows a trend since 1990 (Figure 4.14b) which shows a steady increase in the reported incidence of those 65+ years which seem to have stabilised post 2010.

### 4.6.3 Poisson regression models

Both univariate and multivariate Poisson regression models show that increasing deprivation is associated with decreasing incidence of campylobacteriosis infections. This agrees with the descriptive analysis carried out above and also with the previous geography study in 2000-06 where the Carstairs index was used as an index of deprivation. This pattern has also been found in Denmark (Simonsen, Frisch et al. 2008), Auckland in New Zealand (Spencer, Marshall et al. 2012) and in individuals >9 years old from Connecticut in the USA (Bemis, Marcus et al. 2014). However, in Denmark no difference was found between those in full time work or study compared with those who were unemployed (Kuhn, Nielsen et al. 2018) and no differences in incidence was associated with the social deprivation index in the Canterbury region of New Zealand (Spencer, Marshall et al. 2012).

Increasing human population density was also associated with decreasing incidence. It is worth noting that this was maintained in the multivariate model along with the deprivation risk factor. It was also significant in the 2000-06 geography study. It is plausible, as discussed in the descriptive statistics section above, that those living in higher population density areas have less environmental exposure due to contact with farm animals, and their faeces, and are also less likely to drink from a private water supply.

Higher incidence of disease was associated with the east in both univariate and multivariate regression models. This was despite FF being the health board with lowest incidence. Looking at the map in Figure 4.9 suggests that incidence is higher to the east, though there is considerable heterogeneity. The univariate model indicated that there was increased risk of campylobacteriosis to the south - a finding that became significant in the multivariate model. This trend is not obvious in the map (Figure 4.9).

Increasing sheep density is a risk factor in both univariate and multivariate models. It is known that sheep shed *Campylobacter* and hence can be a source of human infection. However, cattle density is a risk factor in the univariate but is protective in the multivariate, but it is worth noting that there is correlation

between cattle and sheep densities (see Figure A4.1.8 in Annex 7.1). Again it is known that cattle shed *Campylobacter* and thus it would be expected that they would be a risk factor. Poultry density is protective in both univariate and multivariate models. This is at odds with a previous study which showed that there is an increased risk around poultry abattoirs and farms (Rosenberg Goldstein, Cruz-Cano et al. 2016) whereas another study found this to be unimportant (Spencer, Marshall et al. 2012). It may be that most poultry in Scotland are kept in broiler houses and as such exposure of the local population is low as the faecal material is located within the house and disposal is predominantly by incineration rather than spreading on fields (Nick Sparks, SRUC, personal communication). In the previous geography study the multivariate analysis did not find that cattle, sheep and poultry densities were significant (Anon. 2007).

Private water supply density was not significant in the univariate model but became significant in the multivariate model. The plausibility of private water supplies being a vehicle for human campylobacteriosis has been evidenced from outbreaks (G. Nichols, Lane et al. 2009), quantitative microbiological risk assessment (Murphy, Thomas et al. 2016) and also the case-control study carried out by HPS in Aberdeen city and shire during August 2005 to November 2007 which identified it as a risk factor (Anon 2010).

The Poisson regression analysis was performed only on total cases. Further work could be done to carry it out on particular sub-groups of the population. For example, if it was done on <5 year olds, this may enable elucidation of the sources of infection in this age group which are known to have excess cases in rural areas. It should also be noted that this type of analysis relies on the address of the case (in this study only the datazone are available). It is known that many cases of campylobacteriosis infection are travel associated (e.g. 18.5% travel abroad and 47% travel within the UK (Anon 2010)) and hence the postal address may not be where the infection was contracted.

### 4.6.4 Logistic and Multinomial Regression

The logistic regression points to risk factors that discern between the least and most deprived populations. For the following risk factors; gender, human population density, properties on PWS and sheep density, the results from the multivariate analysis were the same as in the univariate analysis. With proportionally fewer female and urban cases in the least deprived areas compared with male and rural cases. It is unclear why there should be this difference - particularly for the female cases. It could partly be due to a higher SIMD1 (most deprived) population being present in urban areas.

In areas with high sheep density and presence of private water supplies there are proportionally more cases resident in least deprived data zones. This could be due to more people living in least deprived areas with private water supplies and higher densities of sheep.

The results from the multivariate analysis were similar to the univariate analysis with respect to age. In the 0-4 and 25-64 year age groups there were proportionally fewer cases living in least deprived areas than for the over 65

year-olds. Health boards GR and LO have proportionally more cases resident in least deprived areas than Tayside (the reference). This may be due to Tayside having a lower proportion of people living in the least deprived quintile. There were inconsistent results for cattle density and poultry density though in the multivariate model lower poultry densities were associated with proportionally more cases in least deprived areas.

In the logistic regression, which compares SIMD5 with SIMD1, if there is a significant difference found then this would be expected to change gradually when comparing SIMD1 with SIMD4, SIMD1 with SIMD3 and SIMD1 with SIMD2. These are the comparisons that the multinomial analysis performs and broadly shows this pattern for gender, longitude and age groups. For all of the other risk factors, except season and health board they tend to show either consistent (or almost consistent) odds ratios >1.0 or <1.0 but the pattern is not a gradient. This suggests that there is at least some consistency when comparing each of the quintiles with the most deprived SIMD1 quintile.

The univariate and multivariate multinomial regression provides broadly similar results. The main difference arises with poultry density (Annex 7.2). In the multivariate analysis, areas with low poultry density have a higher proportion of cases in SIMD5, SIMD4, SIMD3, SIMD2 compares with SIMD1. The opposite is the case in the univariate analysis (Figure 4.12(i)).

### 4.6.5 Proximity of case to GP practices

There were some significant differences found when comparing cases with controls from the general population. However, these differences were small and unlikely to explain the excess of reported cases in the least deprived population in Scotland.

### 4.6.6 Long term variation in reported cases

Campylobacteriosis incidence has increased since 1990. There have been some apparent peaks and troughs in the subsequent 25 years and incidence now appears to be rising. There has been a notable drop in incidence in the <5 years old age group during the last 27 years. NHS 24 which was launched in 2002 may have had some effect on the figures (and possibly the other age groups) but there has been no research that the authors are aware of that has been done to support this hypothesis. Incidence in the 65+ age group increased until 2010 and is now relatively stable.

## 4.7 Conclusions

Human campylobacteriosis is more common than it was in 1990. This partly reflects increase in incidence in adults and the elderly (>65 years). The incidence in young children (<5 years) has fallen throughout the period. There is an excess incidence of 11% in the male population. The reasons for this are unclear but are likely to include both physiological and behavioural factors.

Focussing on deprivation, there remains an excess of 19% of campylobacteriosis cases in the less deprived SIMD quintiles (i.e. SIMD2 to SIMD5). Poisson regression suggests that deprivation is protective in both the univariate and multivariate analysis.

Logistic regression, comparing cases resident in SIMD1 and SIMD5, indicates that proportionally there are fewer female and urban (compared with male and rural) cases in least deprived (SIMD5) data zones. Among cases in areas with high sheep density and private water supplies there are proportionally fewer residents of most deprived (SIMD1) areas. Most of the risk factors used in the multinomial analysis either follow a trend or similar pattern of values for the odds ratio as in the logistic regression.

There is no strong evidence to suggest that living close to a GP increases the likelihood of being reported as a campylobacteriosis case. There is some evidence to suggest that people living in SIMD1 (most deprived) areas within 1km of a GP with a campylobacteriosis infection are less likely to attend their primary healthcare provider (and thus become a recognised 'case') than expected. This is a small effect (2% of cases in SIMD1 areas).

# 5. Hospitalised Case Study

## 5.1 Introduction

This chapter deals with the analysis of the human campylobacteriosis hospital discharge data from Scotland. This involves three approaches.

Section 5.3 utilises both descriptive and analytical (i.e. Poisson, logistic and multinomial regression) approaches to identify risk factors and the pattern of hospital inpatient episodes.

Section 5.4 uses the hospital discharge data together with the locations of hospitals to identify whether hospitalisations are more likely to happen if people live close to a hospital.

Section 5.5 utilises long term (1990-2017) hospital discharge data to identify changes in secular trends.

## 5.2 Overview of data sources

### 5.2.1 Retrospective hospitalisation data

The National Health Service in Scotland collates hospital discharge data on human campylobacteriosis from each hospital through eDRIS (The electronic Data Research and Innovation Service), which is part of ISD (Information Services Division). Data for this study were obtained from eDRIS for two main reasons:

- to determine the proportion of inpatient stays attributable to residents of deprived areas relative to less deprived areas and
- to describe the epidemiology of campylobacteriosis inpatient episodes across Scotland.

Non-patient identifiable hospital discharge data (age, date of admission, length of stay, gender, health board and data zone) for the four years and three months previous to the start of the current study (1st January 2012 to 31st March 2016) were obtained from across Scotland. These comprised 3,806 hospital discharges.

### 5.2.2 Prospective hospitalisation data

Non-identifiable Scottish hospital discharge data (age, date of admission, length of stay, gender, health board and data zone) for the two years from 1st April 2016 to end of March 2018 were obtained. These comprised 1,940 hospital discharges. Since the case-control study ran for an additional 5 months summary hospitalisation data were obtained from eDRIS providing the number of hospital discharges by month by health board. These comprised an additional 607 hospital discharges. These final 5 months of data were supplied at the end of the study and were included only in those analysis where specified.

### 5.2.3 Long term summary hospitalisation data

Summary campylobacteriosis hospital discharge data from Scotland during 1997 to 2011 (n=6,557) were obtained from the literature (N. J. C. Strachan, Rotariu et al. 2013). These enabled long term trends to be determined when combined with the data in 5.2.1 and 5.2.2.

## 5.3 Descriptive and analytical epidemiology of retrospective and prospective campylobacteriosis hospitalisation inpatient episodes

### 5.3.1 Aims

This section aims (i) to provide a description of human campylobacteriosis hospitalisation in Scotland stratified by age, gender, deprivation, health board, rurality and temporal trends and (ii) to identify risk factors for human campylobacteriosis hospitalisations and in particular those factors that may correlate with deprivation.

### 5.3.2 Data

A hospital discharge from campylobacteriosis (recorded by ISD in Scotland) is defined as a person leaving the hospital after being admitted for *Campylobacter* enteritis or being diagnosed as having campylobacteriosis during their stay (ICD-10-CM Diagnosis Code A04.5, https://icd.codes/icd10cm/A045 ) alone or in conjunction with other diagnoses.

Collection of hospitalisation data for 1st January 2012 to 31st March 2018 is described in Chapter 5.2.1.

Non-disease data used in the analysis are described in Chapter 4, section 4.3.2.1.

### 5.3.3 Methods

#### 5.3.3.1 Descriptive epidemiology

Graphs and tables were generated to illustrate how human campylobacteriosis hospitalisation stays vary with age, gender, deprivation, health board, rurality and time. Summary statistics of the length of hospitalisation was also determined.

SPSS Statistics v24 was utilised to determine the difference in incidence of hospitalisation between health boards by Analysis of Variance using Tukey's honest significant difference (Tukey 1949) with post-hoc correction (Bonferroni) for multiple comparisons. The student's t-test was used to compare mean incidence values between groups (Clifford-Blair, Higgins 1980).

To visualise the number of hospitalisations and incidence of human campylobacteriosis hospital discharges, maps were produced in ArcMap 10.5 (http://www.arcgis.com ).

#### 5.3.3.2 Univariate and multivariate Poisson regression

Univariate and multivariate Poisson regression analysis (Gardner, Mulvey et al. 1995) was performed on the hospital discharge data (1st January 2012 to 31st March 2018), using SPSS Statistics v24, as for cases of human campylobacteriosis (see Chapter 4, section 4.3.2.3).

#### 5.3.3.3 Univariate and multivariate binary logistic regression

Univariate and multivariate binary logistic regression (Cox 1958, Kleinbaum, Klein 2010) analysis looked for differences between risk factors for hospitalisation from the *most* and *least* deprived quintiles. The method was similar to that described in Chapter 4, section 4.3.2.4.

#### 5.3.3.4 Univariate and multivariate multinomial logistic regression

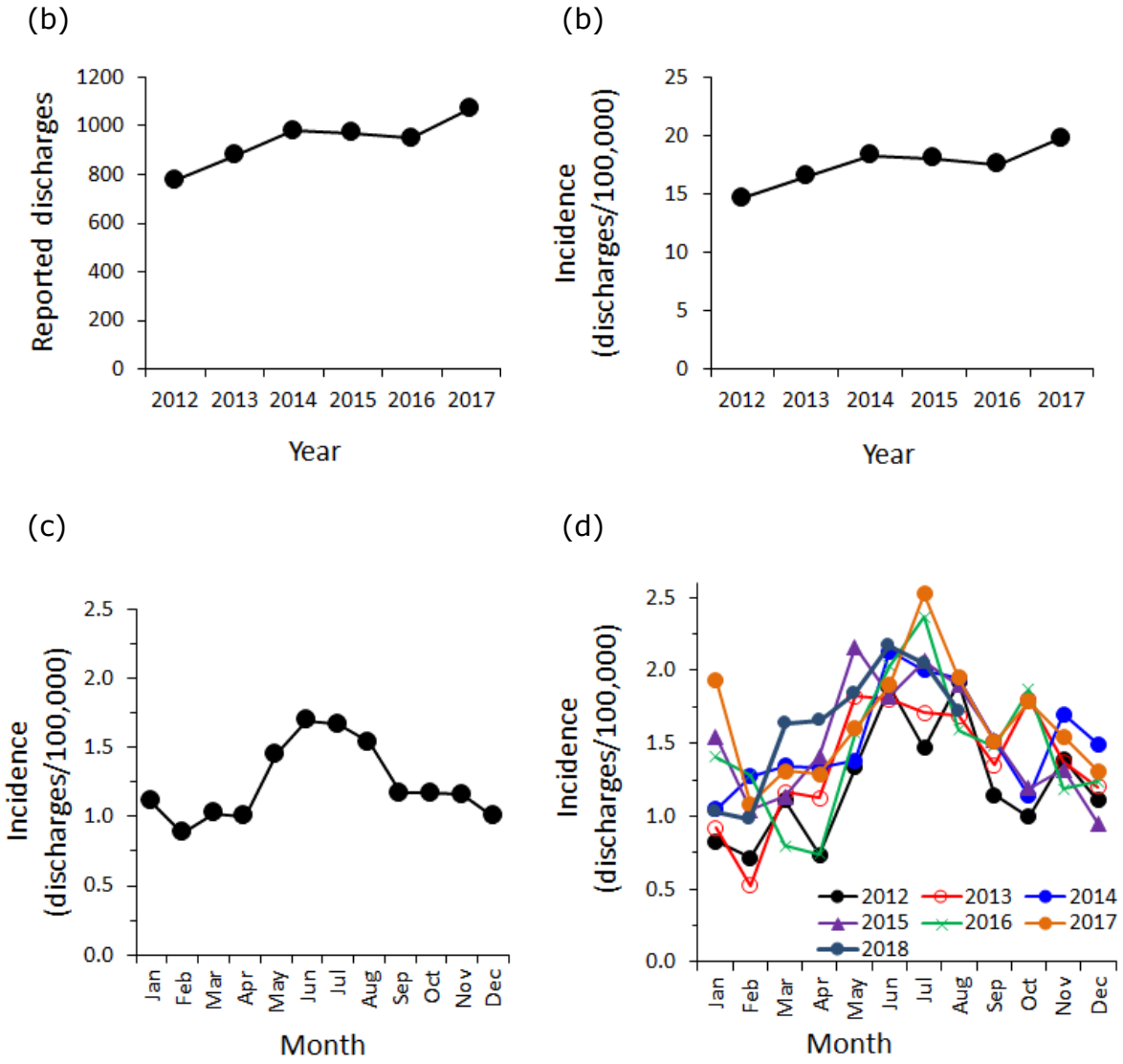
Univariate and multivariate multinomial logistic regression (Varga, Middleton et al. 2012) was used to look for differences between risk factors for hospitalisation from all 5 SIMD quintiles. The method was described in the human campylobacteriosis case analysis (see Chapter 4, section 4.3.2.5).

### 5.3.4 Results and Discussion

#### 5.3.4.1 The epidemiology of human campylobacteriosis hospitalisation in Scotland

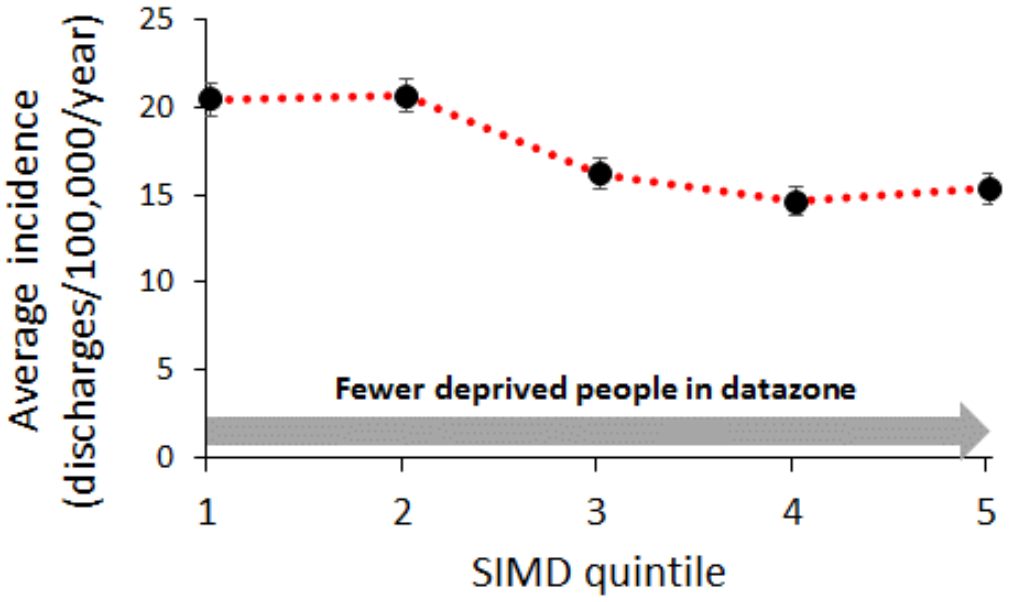
In total 6,353 hospitalisations for campylobacteriosis were reported (this includes all hospital admissions even if there was not an overnight stay). Of these, 5,082 (80%) had campylobacteriosis as the main diagnosis. This corresponds to 15.5% of reported cases. Human campylobacteriosis hospitalisation had an increasing trend during 2012 to 2017 (Figure 5.1(a) and (b)). There is a "summer" peak that occurs between May and August (Figure 5.1(c)). The summer incidence of 1.6 ±0.1 discharges/100,000/month was significantly ($P=6.9\times10^{-5}$) higher than for the rest of the year (1.22±0.14 discharges /100,000/month). A second peak, of much smaller size, appears to have occurred in October 2013, 2016 and 2017 but not in the other years (Figure 5.1(d)). The incidence of campylobacteriosis hospitalisation during 1st January 2012 to 31st March 2018 was 17.3±1.4 hospital discharges/100,000. There were no published data for comparison in the previous geography (S14004) study 2000-2006 but data obtained by the authors from ISD during this time period indicates that there is no trend by deprivation quintile. It should however be noted that this used the Carstairs index as a measure of deprivation rather than SIMD.

Figure 5.1. Hospitalisation and incidence of human campylobacteriosis hospitalisation in Scotland




(a) Hospitalisation and (b) incidence of human campylobacteriosis hospitalisation in Scotland Jan 2012 -Dec 2017, (c) monthly incidence in Scotland between 1st Jan 2012 – 31st Aug 2018 and (d) monthly incidence by year 1st Jan 2012 – 31st Aug 2018.

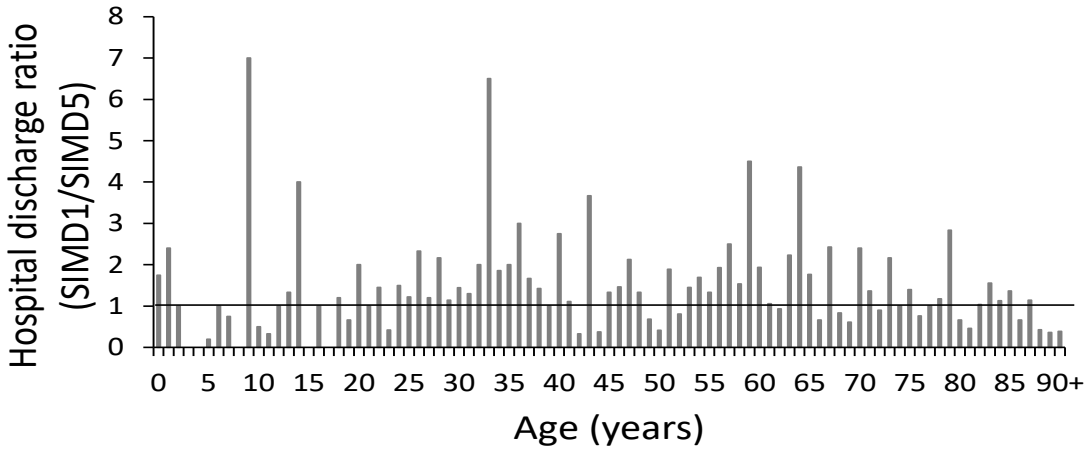
**Figure 5.2. Variation in incidence of hospitalisation by SIMD Quintile (2012-2017).**




Variation in incidence of hospitalisation by SIMD Quintile (2012-2017). The largest number of deprived people live in SIMD quintile 1 whilst the fewest live in SIMD quintile 5.

There was an excess of hospitalisation (9.2%) in the first two SIMD quintiles (more deprived) compared with the three less deprived SIMD quintiles (Figure 5.2). Also, there was a significantly (P=0.028) higher rate of hospitalisation among those from most deprived than least deprived data zones on average across all ages (Figure 5.3).

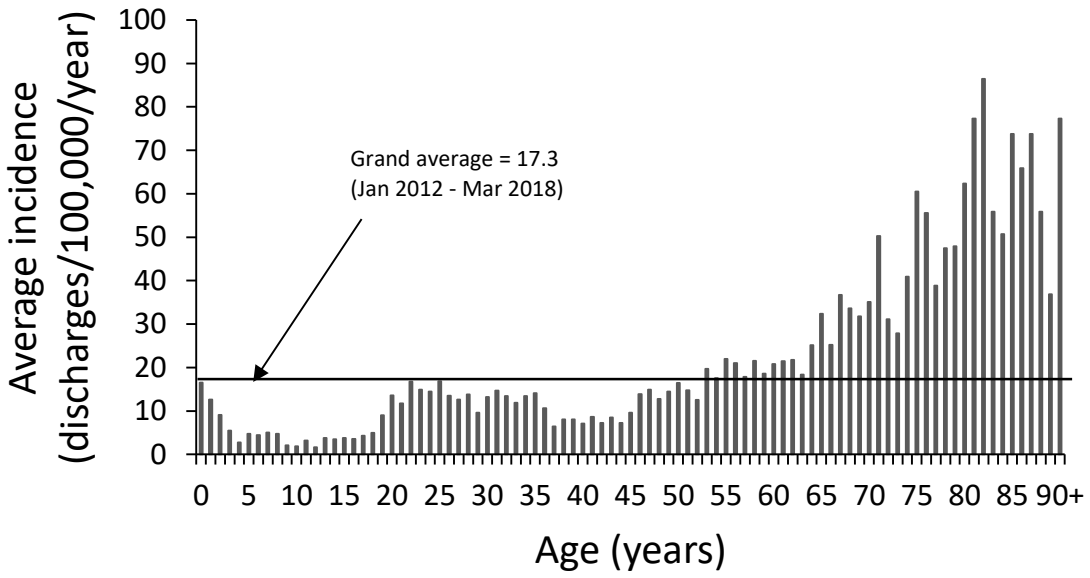
**Figure 5.3. Ratio of most to least deprived hospitalisation by age**




The average incidence of hospitalisation across the 6.25 year period of this study was stratified by age (Figure 5.4). The incidence in those over 65 years of age (43.4±4.2 hospital discharges/100,000/year) was significantly (P<0.05) higher than the average (17.3±1.4 discharges/100,000/year) (Figure 5.4). Also,

incidence in those <50 years of age (9.4±1.3 hospital discharges/100,000/year) was significantly (P<0.05) lower than the average (Figure 5.4).

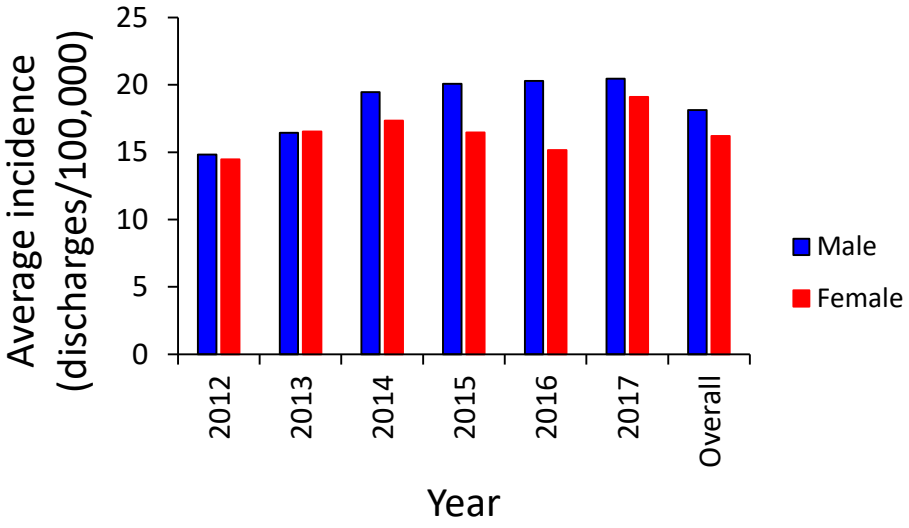
**Figure 5.4. The average incidence of campylobacteriosis hospitalisation in Scotland by age**




1st Jan 2012 -31st Mar 2018

Figure 5.5 shows the yearly campylobacteriosis hospitalisation rate by gender. Although it appears to be a higher incidence in males this is not statistically significant (P=0.054).

**Figure 5.5. The average rate of hospitalisation with campylobacteriosis hospitalisation in Scotland by gender and year**

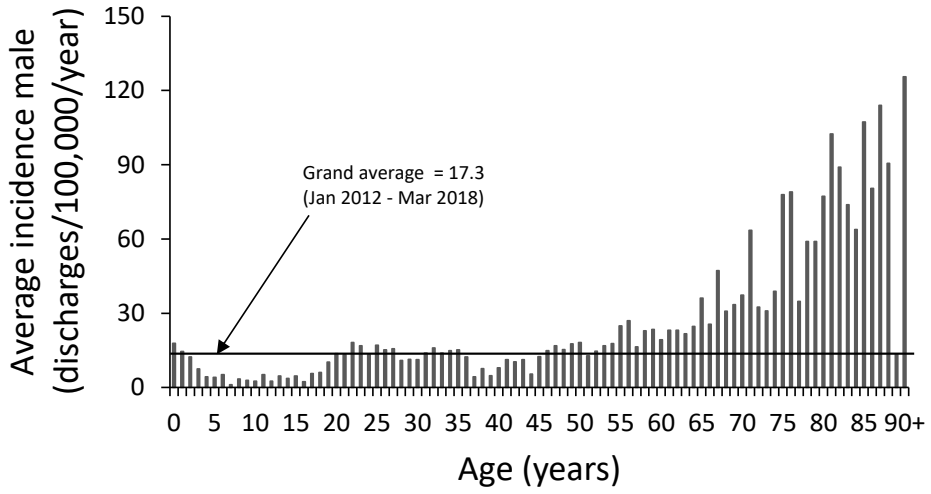



2012-2017

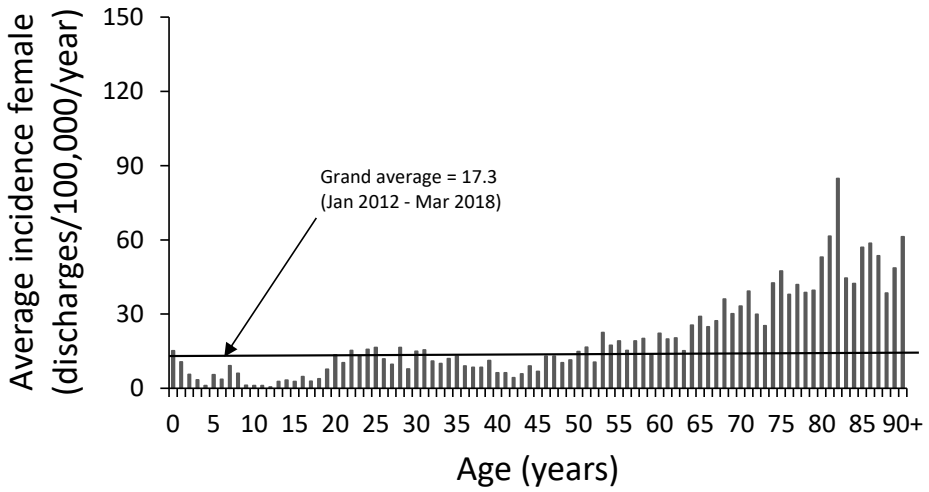
When the average incidence of campylobacteriosis hospitalisation was stratified by age for each gender (Figure 5.6 (c)), it was shown that the incidence across ages was, on average, higher in males than in females (incidence ratio >1, P=0.028). Also, the incidence of hospitalisation is higher (P<0.0001) than average for those >65 years old for both males and females (Figure 5.6(a) and (b)).

**Figure 5.6. The average incidence of campylobacteriosis hospitalisation by gender**

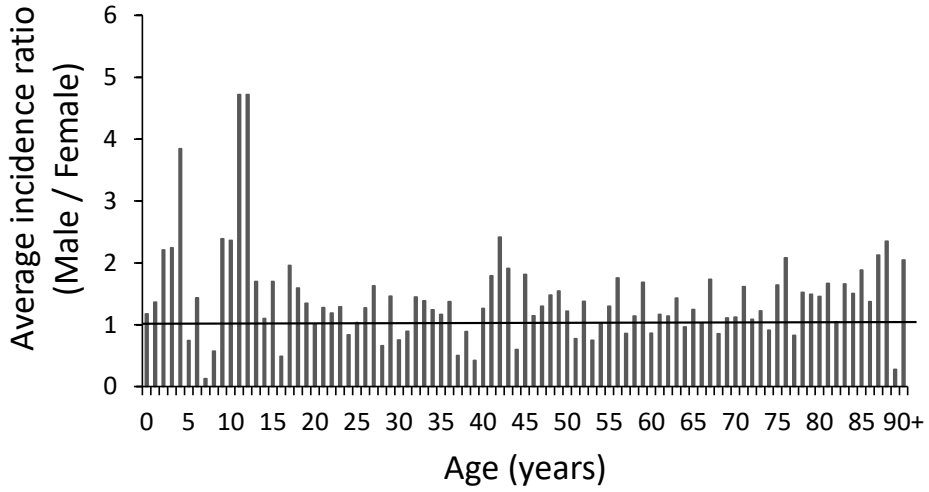
(a)




(b)




(c)




in (a) male, (b) female and the (c) male:female incidence ratio stratified by age.

Examination of rates of human campylobacteriosis hospitalisation across mainland health boards reveals that GC had the highest incidence (22.0 discharges/100,000 people) (Table 5.1 and Figure 5.8) and GR the lowest (11.8

discharges /100,000 people). The incidence in GR appears to be decreasing during the study period, whilst for GC the trend is upwards (Figure 5.7). The analysis of variance performed to determine whether there are differences in incidence of hospitalisation between each of the mainland health boards shows that the incidence in BR, GR and HG was significantly lower ($P<0.05$) than in GC (see Table 5.2 and Figure 5.8).

**Figure 5.7. Incidence of human campylobacteriosis hospitalisation stratified by year for mainland health boards in Scotland.**

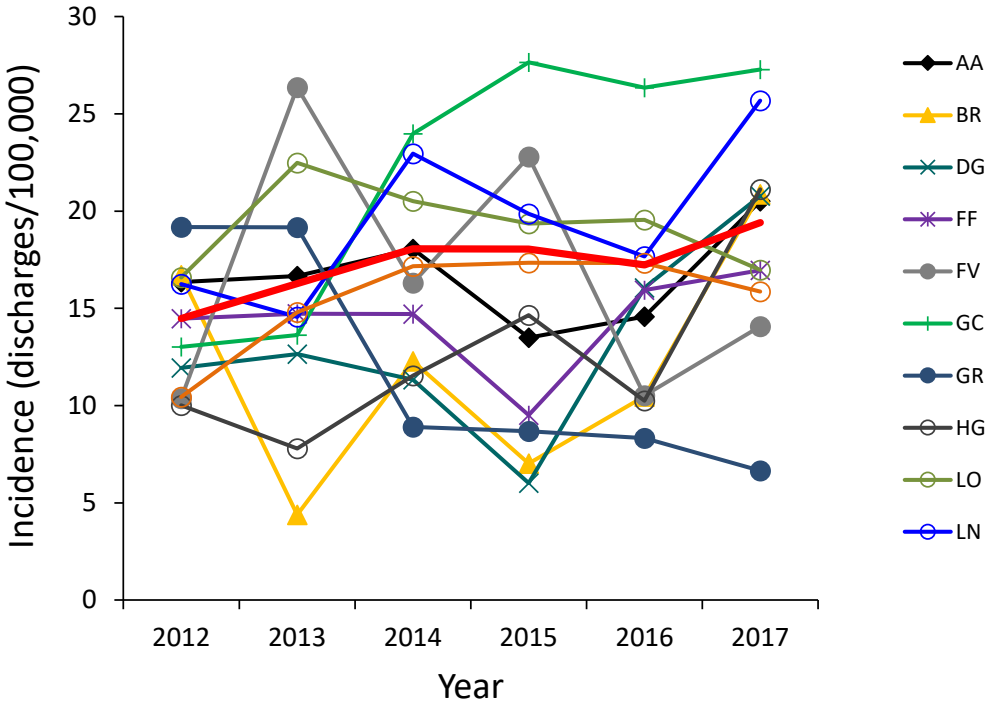
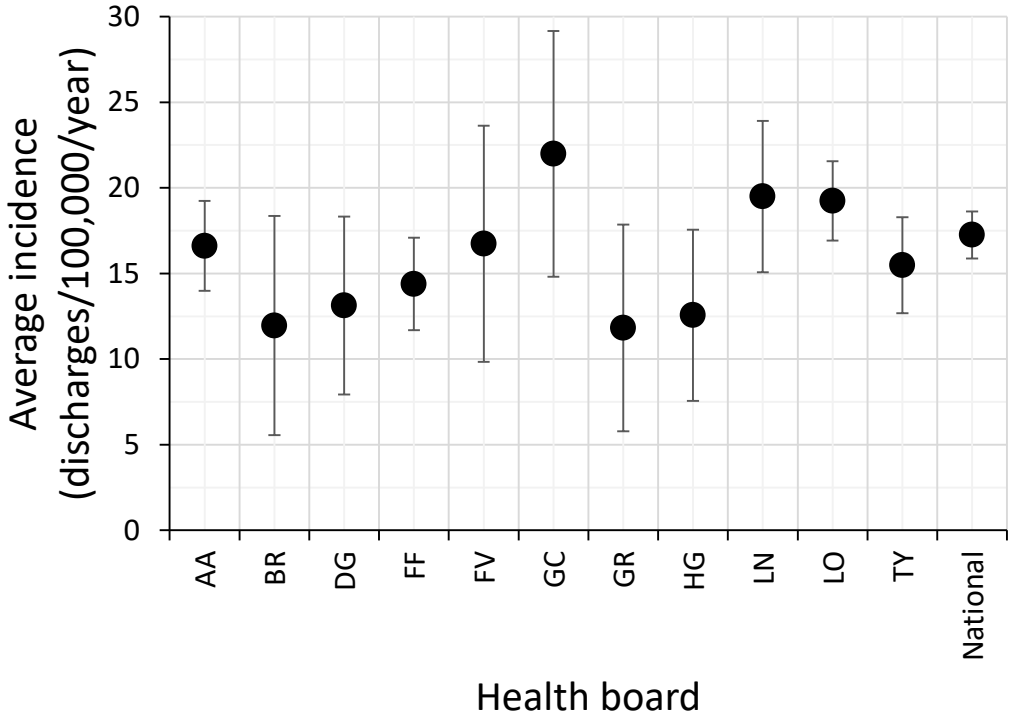


**Table 5.1. Campylobacteriosis hospitalisation incidence by health board.**

| Health board | Average incidence (discharges/100,000/year) (Jan2012 - Mar2018) |
|---|---|
| Ayrshire & Arran (AA) | 16.6 |
| Borders (BR) | 12.0 |
| Dumfries & Galloway (DG) | 13.1 |
| Fife (FF) | 14.4 |
| Forth Valley (FV) | 16.7 |
| Greater Glasgow & Clyde (GC) | 22.0 |
| Grampian (GR) | 11.8 |
| Highland (HG) | 12.6 |
| Lanarkshire (LN) | 19.5 |
| Lothian (LO) | 19.2 |
| Tayside (TY) | 15.5 |
| Orkney (OR) | 13.8 |
| Shetland (SH) | 7.2 |
| Western Isles (WI) | 8.5 |
| National | 17.3 |

**Figure 5.8. The average incidence of hospitalisation by health board**




Error bars represent 95% Cis.

**Table 5.2. Difference in incidence of hospitalisation between health boards by Analysis of Variance**

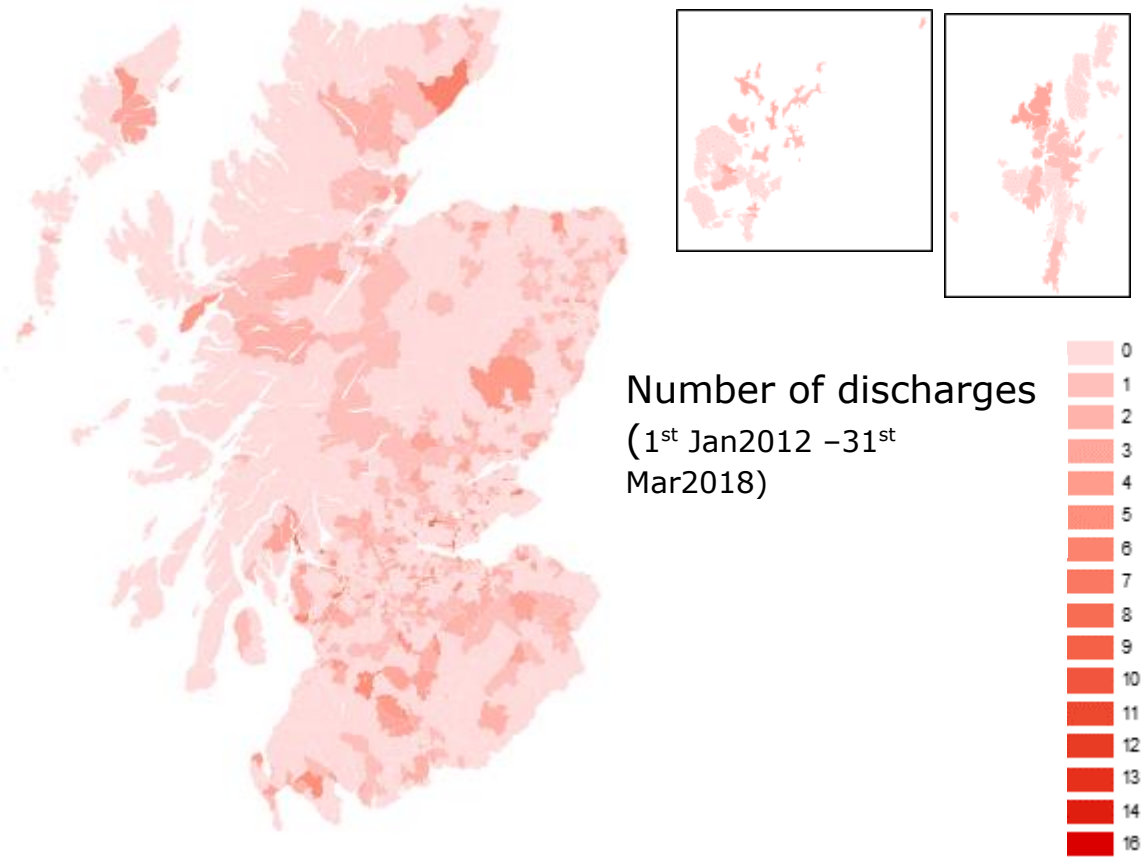
| Contrast | Difference in incidence between health boards (discharges/100,000 people) | P-value (of seeing observed difference or greater) |
|---|---|---|
| BR<GC | -10.0 | 0.023 |
| GR<GC | -10.2 | 0.020 |
| HG<GC | -9.4 | 0.042 |

Analysis of Variance using Tukey's honest significant difference with post-hoc correction (Bonferroni) for multiple comparisons.

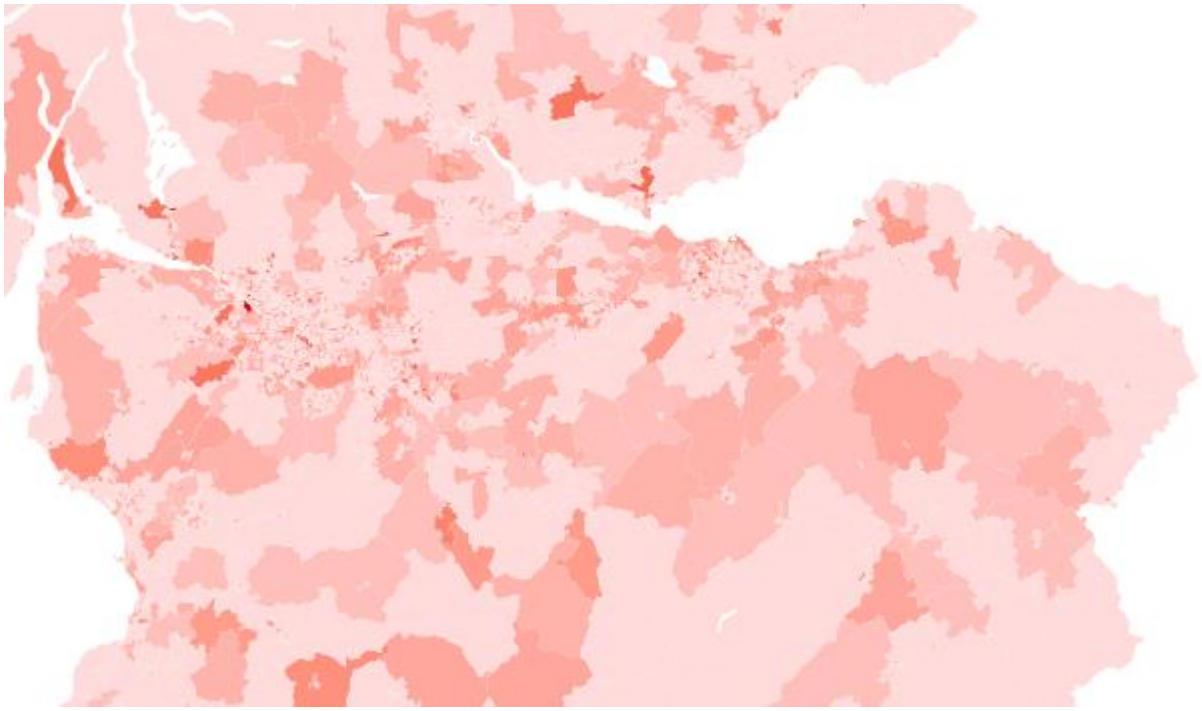
Both the number of hospitalisations (Figure 5.9(a) and (b)) and corresponding incidence (Figure 5.9(c) and (d)) appear to be heterogeneous across Scotland. This is in part due to statistical fluctuation as there are 6,505 data zones in Scotland with on average only 0.9 hospitalisations in each during the study period.

**Figure 5.9. Numbers and incidence of human campylobacteriosis hospitalisation for SIMD data zones**

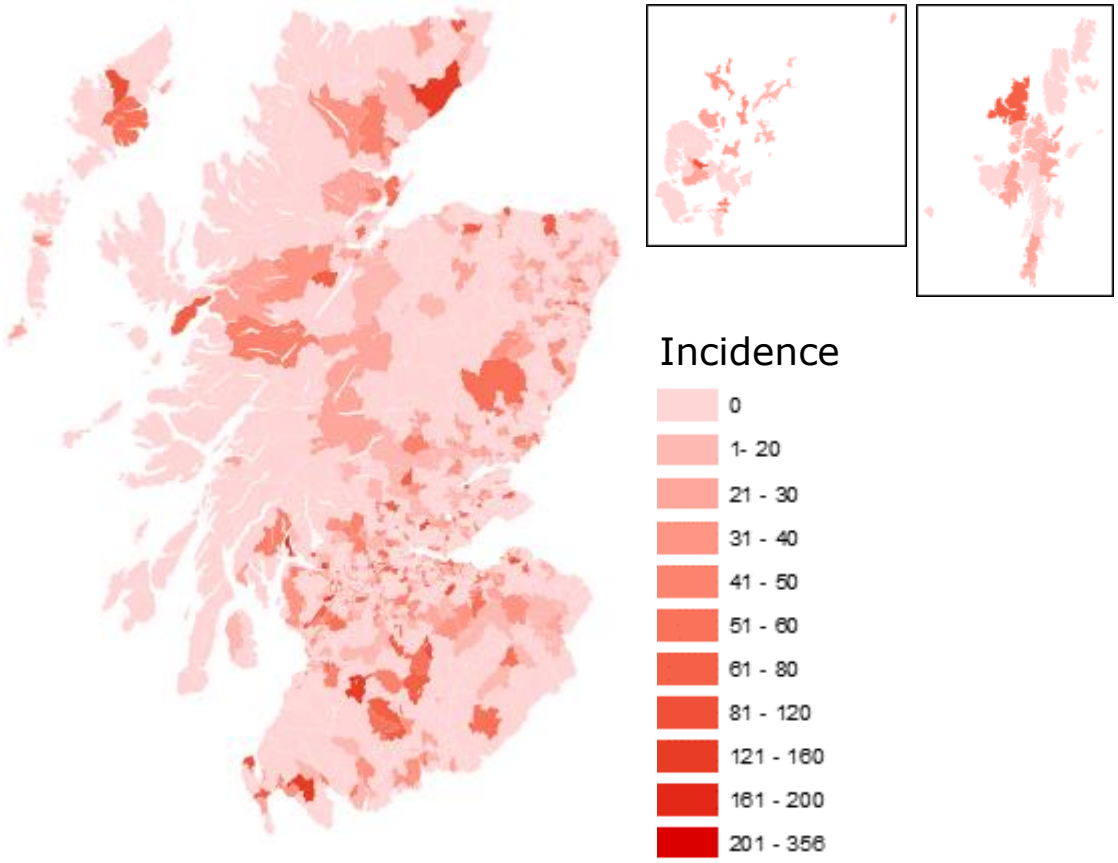
(a)




(b)



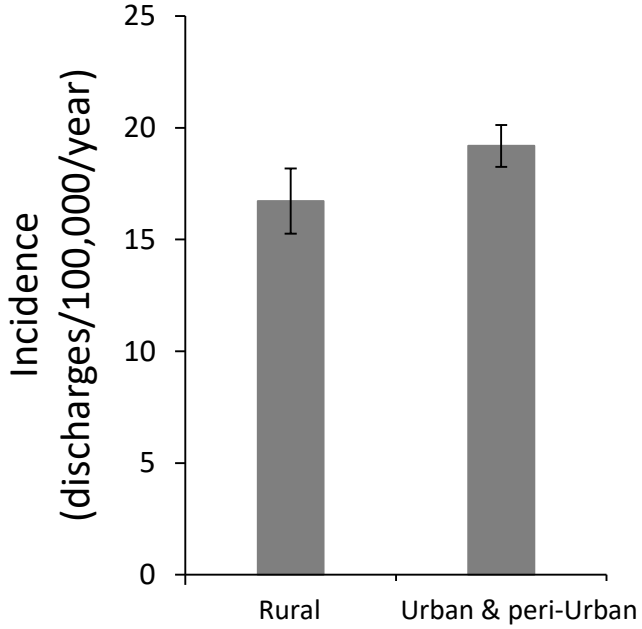
(c)




(d)

Numbers (a) and (b), and incidence (c) and (d) of human campylobacteriosis hospitalisation for SIMD data zones (1st Jan2012 –31st Mar2018).

**Figure 5.10. Hospitalisation incidence of rural and urban/peri-urban populations**




Average incidence and 95% CIs were calculated at data zone level. (Threshold population density: Rural ≤200 people/km$^2$; Urban and peri-Urban >200 people/km$^2$).

The incidence of campylobacteriosis hospitalisation in urban and peri-Urban data zones (Figure 5.10) was significantly higher than in rural data zones (P=0.003). This excess comprises 14.5% of the total hospitalisation in Scotland.

**Figure 5.11. Frequency of the duration of hospitalisation (nights).**

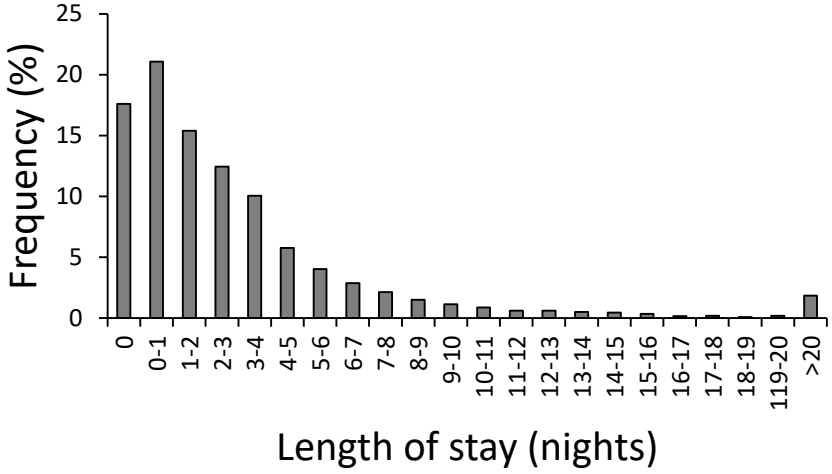



**Table 5.3. Summary statistics of duration of hospitalisation (nights).**

| Summary statistics | Length of stay (nights) |
|---|---|
| Mean | 3.74 |
| Standard Error | 0.09 |
| Median | 2 |
| Mode | 1 |
| Standard Deviation | 6.62 |
| Minimum | 0 |
| Maximum | 166 |

There were 5478 hospitalisation discharges in total.

The distribution of hospital length of stay (in nights) is left skewed with a long tail (Figure 5.11, Table 5.3). Eighteen percent (18%) of people admitted to hospital did not stay overnight. The modal length of stay was one night (21%) and 15% stayed for the median two nights. The mean stay was 3.7 nights with a maximum of 166 and <2% staying for more than 20 nights.

### 5.3.4.2 Risk factors associated with human campylobacteriosis hospitalisation in Scotland. Results from univariate and multivariate Poisson regression

The univariate Poisson regression analysis (Table 5.4) shows that increasing human population density and deprivation (by SIMD score) were positively associated ($P<0.05$) with increasing campylobacteriosis hospitalisation rates. Increasing cattle, sheep, poultry and PWS densities were associated ($P<0.05$) with decreasing incidence of hospital discharges. Longitude was negatively associated with incidence of hospitalisation (i.e. lower incidence towards the east). Latitude was negatively associated with incidence of hospitalisation (i.e. lower incidence towards the north). This follows the population as most of the population live in SW Scotland (Glasgow and Lanarkshire etc.).

**Table 5.4. Univariate Poisson regression analysis of risk factors for campylobacteriosis hospital discharges**

| Variable (risk factor) | Unit | Estimate of regression coefficient ($\beta$) | Std. Error | P-value |
|---|---|---|---|---|
| Latitude | degree | -0.178 | 0.0214 | <0.001 |
| Longitude | degree | -0.074 | 0.0164 | <0.001 |
| SIMDScore* | - | 0.008 | 0.0008 | <0.001 |
| Human population density | people/km$^2$ | $6.93\times10^{-6}$ | $3.13\times10^{-6}$ | 0.027 |
| Private water density | Number of properties / number of people | -5.237 | 0.6944 | <0.001 |
| Poultry density | poultry/km$^2$ | $-2.35\times10^{-5}$ | $7.82\times10^{-6}$ | 0.003 |
| Cattle density | cattle/km$^2$ | -0.003 | 0.0005 | <0.001 |
| Sheep density | sheep/km$^2$ | -0.002 | 0.0002 | <0.001 |

If the regression coefficient ($\beta$) is positive the incidence increases as the risk factor increases and if it is negative it decreases (for brevity the intercepts are not provided). The p-values indicate statistical significance. The red colour indicates a significant increase in incidence when the risk factor increases, whilst blue is the opposite showing a significant decrease and black shows no significant difference. *As deprivation increases then the incidence of campylobacteriosis hospitalisation increases.

The multivariate Poisson regression analysis (Table 5.5) shows that increasing deprivation is positively associated (P<0.05) with increasing incidence of human campylobacteriosis hospitalisation (as in univariate analysis). Increasing, latitude (i.e. further north), human population density, PWS density, poultry density, cattle density and sheep density was associated with decreasing incidence of hospitalisation (i.e. protective for hospitalisation). Longitude (west to east) was no longer significant.

**Table 5.5. Multivariate Poisson regression analysis of risk factors for campylobacteriosis hospital discharges**

| Variable (risk factor) | Unit | Estimate of regression coefficient ($\beta$) | Std. Error | P-value |
|---|---|---|---|---|
| Intercept | - | 3.697 | 1.4853 | 0.013 |
| Latitude | degree | -0.185 | 0.0258 | <0.001 |
| Longitude | degree | 0.008 | 0.0196 | 0.694 |
| SIMDScore* | - | 0.006 | 0.0008 | <0.001 |
| Human population density | people/km$^2$ | $-1.74\times10^{-5}$ | $3.82\times10^{-6}$ | <0.001 |
| Private water density | Number of properties / number of people | -4.027 | 0.6832 | <0.001 |
| Poultry density | poultry/km$^2$ | $-2.47\times10^{-5}$ | $7.77\times10^{-6}$ | 0.001 |
| Cattle density | cattle/km$^2$ | -0.003 | 0.0005 | <0.001 |
| Sheep density | sheep/km$^2$ | -0.001 | 0.0003 | 0.001 |

If the regression coefficient ($\beta$) is positive the incidence of hospitalisation increases as the risk factor increases and if it is negative it decreases. The p-values indicate statistical significance. The red colour indicates a significant increase in incidence when the risk factor increases, whilst blue is the opposite showing a significant decrease and black shows no significant difference. *As deprivation increases then the incidence of campylobacteriosis hospitalisation increases.

### 5.3.4.3 Risk factors associated with campylobacteriosis hospitalisation in Scotland. Results from univariate and multivariate binary logistic regression

**(i) <u>Univariate logistic regression</u>**

Table 5.6 presents the results from the univariate binary logistic regression comparing the least (SIMD5) and most (SIMD1) deprived quintiles for each risk factor. Statistically significant differences (P <0.05) are colour coded as follows: a proportional decrease in the least deprived number of hospital episodes is coloured in blue, whilst an increase is coloured in red. Black colour means that the result is not statistically significant. The interpretation of the following results is done in a similar way as those presented in Chapter 4, Section 4.3.3.3 for the analysis of reported cases.

*Gender:* the ratio of the number of 'least deprived' to number of 'most deprived' amongst hospital discharges, did not vary by gender.

*Human population density*: for patients resident in least deprived areas, the number of hospital episodes in the peri-urban and rural populations is proportionally higher than that in the urban (*reference*) population.

*Longitude*: the ratio of the number of inpatients from least deprived to those from most deprived areas, increases by a factor of 2.208 towards the East of Scotland for each degree increase in longitude (a degree corresponds to

approximately 50 miles in Scotland). Hence, there are proportionally more inpatients from least deprived areas in the east compared with the west.

*Latitude*: the ratio of the number of inpatients from least deprived to those from most deprived areas, in the north is higher than in the south (i.e. there are proportionally more inpatients from least deprived areas in the north).

*Age*: the ratio of the number of inpatients from least deprived to those from most deprived areas, in 25-64 years old is significantly lower than in 65+ years old population (the reference group). This means that in the 25-64 years old there are proportionally fewer inpatients from least deprived areas than in the 65+ years old population. There were no other significant differences.

*PWS's*: the ratio of the number of inpatients from least deprived to those from most deprived areas, is higher in data zones where there are PWS's. This means that there are proportionally more inpatients from least deprived areas in data zones where there are properties on PWS's.

*Cattle density*: the ratio of the number of inpatients from least deprived to those from most deprived areas, in data zones with "Mid1" (intermediate) cattle density is lower than in data zones with high cattle density. This means that the number of inpatients from least deprived areas in data zones with "Mid1" cattle density is proportionally lower than in data zones with high cattle density. There were no other significant differences. Hence, the results appear to be inconsistent as it would be expected that "Low" cattle density would also be significant.

*Sheep density*: there is a decreasing trend in the ratio of the number of inpatients from least deprived to those from most deprived areas, as sheep density decreases. So in higher sheep densities there are proportionally more inpatients from least deprived areas.

*Poultry density*: the ratio of the number of inpatients from least deprived to those from most deprived areas, did not vary by poultry density.

*Time of year*: the ratio of the number of inpatients from least deprived to those from most deprived areas, in "Summer" is higher than in the rest of year. So in "Summer" there are proportionally more inpatients from least deprived areas.

*Health board*: Greater Glasgow & Clyde health board had the highest incidence of hospitalisation (22.0 hospital discharges/100,000 (95% CI – 14.8 – 29.2)) during the time period between 1st January 2012 to 31st March 2018 and was used as the reference in the logistic regression analysis when comparing health boards.

The ratio of the number of inpatients from least deprived to those from most deprived areas, in DG, FF, FV, GR, HG, LO and TY is higher than in Greater Glasgow & Clyde. This means that in these health boards there are proportionally higher numbers of inpatients from least deprived areas than in Greater Glasgow & Clyde. The situation is opposite for LN, where there are proportionally lower numbers of inpatients from least deprived areas than in Greater Glasgow & Clyde. The ratio of the number of inpatients from least deprived areas to the number of inpatients from most deprived areas for AA and BR was not significantly different from that in Greater Glasgow & Clyde.

## Table 5.6. Univariate binary logistic regression comparing hospitalisation in least deprived and most deprived data zones

| Risk factor | Estimate of regression coefficient ($\beta$) | Std. Error | OR(95% CI) | P-value |
|---|---|---|---|---|
| **GENDER** | | | | |
| **Male (reference)** | | | | |
| Female | -.083 | .084 | .921(.781, .951) | 1.085 |
| **POPULATION DENSITY** | | | | |
| **Urban - High population density (reference)** | | | | |
| peri-Urban - Intermediate population density | .747 | .103 | 2.110(1.726, 2.580) | <0.001 |
| Rural - Low population density | 3.089 | .520 | 21.96(7.93, 60.81) | <0.001 |
| **POSITION (continuous variable)** | | | | |
| Longitude | .792 | .064 | 2.208(1.948, 2.502) | <0.001 |
| Latitude | .903 | .104 | 2.466(2.012, 3.022) | <0.001 |
| **AGE** | | | | |
| **65+ years old (reference)** | | | | |
| 0-4 years old | -.321 | .280 | .726(.419, 1.257) | .253 |
| 5-24 years old | .077 | .152 | 1.080(.802, 1.454) | .613 |
| 25-64 years old | -.454 | .090 | .635(.532, .757) | <0.001 |
| **PRIVATE WATER SUPPLY** | | | | |
| **Properties on PWS (No) (reference)** | | | | |
| Properties on PWS (Yes) | 2.470 | .375 | 11.82(5.67, 24.65) | <0.001 |
| **CATTLE DENSITY** | | | | |
| **Cattle density-High (reference)** | | | | |
| Cattle density-Low | .016 | .123 | 1.016(.799, 1.292) | .899 |
| Cattle density-Mid1 | -.641 | .129 | .527(.410, .678) | <0.001 |
| Cattle density-Mid2 | .023 | .132 | 1.023(.790, 1.326) | .861 |

(continued)

| Risk factor | Estimate of regression coefficient ($\beta$) | Std. Error | OR(95% CIs) | P-value |
|---|---|---|---|---|
| **SHEEP DENSITY** | | | | |
| **Sheep density-High (reference)** | | | | |
| Sheep density-Low | -.501 | .127 | .606(.472, .777) | <0.001 |
| Sheep density-Mid1 | -.486 | .133 | .615(.474, .799) | <0.001 |
| Sheep density-Mid2 | -.103 | .140 | .902(.686, 1.187) | .462 |
| **POULTRY DENSITY** | | | | |
| **Poultry density-High (reference)** | | | | |
| Poultry density-Low | -.151 | .121 | .860(.679, 1.089) | .211 |
| Poultry density-Mid1 | -.119 | .132 | .888(.615, 1.150) | .368 |
| Poultry density-Mid2 | -.227 | .132 | .797(.926, 1.033) | .086 |
| **TIME OF YEAR** | | | | |
| **Rest of year (reference)** | | | | |
| Summer (May, Jun, Jul, Aug) | .193 | .085 | 1.213(1.026, 1.434) | .024 |

(continued)

| Risk factor | Estimate of regression coefficient ($\beta$) | Std. Error | OR(95% CIs) | P-value |
|---|---|---|---|---|
| **HEALTH BOARD** | | | | |
| **GC(reference)** | | | | |
| AA | -.123 | .187 | .885(.613, 1.276) | .512 |
| BR | -.827 | .785 | .437(.094, 2.038) | .292 |
| DG | 2.286 | .637 | 9.84(2.83, 34.27) | <0.001 |
| FF | .488 | .200 | 1.630(1.102, 2.410) | .015 |
| FV | .638 | .209 | 1.893(1.256, 2.855) | .002 |
| GR | 2.028 | .201 | 7.59(5.12, 11.27) | <0.001 |
| HG | .900 | .309 | 2.460(1.343, 4.506) | .004 |
| LN | -.365 | .155 | .694(.512, .940) | .018 |
| LO | 1.345 | .126 | 3.839(2.997, 4.919) | <0.001 |
| TY | .561 | .186 | 1.752(1.218, 2.521) | .003 |

Univariate binary logistic regression comparing hospitalisation in least deprived and most deprived data zones (for brevity the intercepts are not provided). Statistically significant results are coloured in red (comparison is significantly higher than the reference) and blue (comparison is significantly lower).

## (ii) Multivariate logistic regression

All factors having a P-value <0.25 in the univariate analysis were introduced into the multivariate analysis simultaneously. Table 5.7 provides the results and the method of interpretation is similar to the univariate analysis.

## Table 5.7. Multivariate binary logistic regression comparing hospitalisation in the least deprived and most deprived data zones

(Note: the factors with OR >1.0 (i.e. fewer cases from more deprived, compared to less disadvantaged, areas) *and* 95% CI above 1.0 are in red font.

| Risk factor | Estimate of regression coefficient ($\beta$) | Std. Error | OR(95% CI) | P-value |
|---|---|---|---|---|
| **POPULATION DENSITY** | | | | |
| **Urban - High population density (reference)** | | | | |
| peri-Urban - Intermediate population density | .966 | .119 | 2.627(2.08, 3.32) | <0.001 |
| Rural - Low population density | 2.613 | .600 | 13.639(4.21, 44.17) | <0.001 |
| **POSITION (continuous variable)** | | | | |
| Longitude | .362 | .273 | 1.436(.842, 2.450) | .184 |
| Latitude | -.737 | .377 | .478(.228, 1.003) | .051 |
| **AGE** | | | | |
| **65+ years old (reference)** | | | | |
| 0-4 years old | -.957 | .348 | .384(.194, .759) | .006 |
| 5-24 years old | .163 | .172 | 1.177(.840, 1.649) | .343 |
| 25-64 years old | -.477 | .102 | .621(.508, .759) | <0.001 |
| **PRIVATE WATER SUPPLY** | | | | |
| **Properties on PWS (No) (reference)** | | | | |
| Properties on PWS (Yes) | .297 | .462 | 1.346(.544, 3.331) | .520 |
| **CATTLE DENSITY** | | | | |
| **Cattle density-High (reference)** | | | | |
| Cattle density-Low | .300 | .250 | 1.350(.827, 2.203) | .230 |
| Cattle density-Mid1 | -.650 | .197 | .522(.355, .768) | .001 |
| Cattle density-Mid2 | .005 | .182 | 1.005(.703, 1.435) | .979 |
| **SHEEP DENSITY** | | | | |
| **Sheep density-High (reference)** | | | | |
| Sheep density-Low | -1.092 | .235 | .336(.212, .532) | <0.001 |
| Sheep density-Mid1 | -.303 | .177 | .738(.522, 1.045) | .087 |

| | | | | |
|---|---|---|---|---|
| Sheep density-Mid2 | .089 | .164 | 1.094(.792, 1.509) | .587 |

(continued)

| Risk factor | Estimate of regression coefficient ($\beta$) | Std. Error | OR(95% CIs) | P-value |
|---|---|---|---|---|
| **TIME OF YEAR** | | | | |
| **Rest of year (reference)** | | | | |
| Summer (May, Jun, Jul, Aug) | .142 | .098 | 1.153(.952, 1.396) | .146 |
| **HEALTH BOARD** | | | | |
| **Glasgow and Clyde (reference)** | | | | |
| Ayrshire and Arran | -.886 | .260 | .412(.248, .686) | .001 |
| Borders | -2.311 | .943 | .099(.016, .630) | .014 |
| Dumfries and Galloway | .776 | .782 | 2.172(.469, 10.048) | .321 |
| Fife | -.203 | .394 | .816(.377, 1.768) | .607 |
| Forth Valley | .033 | .289 | 1.033(.587, 1.821) | .909 |
| Grampian | 2.298 | .747 | 9.956(2.305, 43.01) | .002 |
| Highland | 1.506 | .579 | 4.510(1.450, 14.032) | .009 |
| Lanarkshire | -.936 | .204 | .392(.263, .585) | <0.001 |
| Lothian | 1.013 | .334 | 2.755(1.431, 5.303) | .002 |
| Tayside | .648 | .458 | 1.912(.779, 4.694) | .157 |
| Intercept** | 42.649 | 21.26 | $3.33\times10^{18}$(na* na*) | .045 |

References are as in univariate and indicated in the table. Statistically significant results are coloured in red (comparison is significantly higher than the reference) and blue (comparison is significantly lower).

* na – not applicable

**The intercept in the logistic regression sets the "baseline" event rate, i.e. the natural logarithm of the odds ratio when all risk factors values are set equal to zero simultaneously (http://www.med.mcgill.ca/epidemiology/joseph/courses/EPIB-621/logistic2.pdf). In practice when there are more than two risk factors (covariates) it is unlikely to have them all set at zero simultaneously. Hence in the above multivariate logistic regression the intercept has no physical meaning. However, using an intercept in the logistic regression is important, otherwise the model will be forced through the origin.

The main findings from Table 5.7 are:

The risk factors *gender* and *poultry density* were removed from the multivariate analysis because they were not significant in the univariate regression analysis.

*Human population density* and *cattle density*: the results from the multivariate analysis were the same as in the univariate analysis.

*Age*: the ratio of the number of inpatients from least deprived to number from most deprived areas, in the 0-4 years old and 25-64 years old groups is significantly lower than for 65+ years old patients (the reference group). This means that for 0-4 and 25-64 years old age groups there are proportionally more inpatients from most deprived areas than in the 65+ years old population. There were no other significant differences. In the univariate analysis only the result for 25-64 years old was significant.

*Sheep density*: the ratio of the number of inpatients from least deprived to those from most deprived areas, in data zones with "Low" sheep density is lower than in data zones with high sheep density. This means that the number of inpatients from least deprived data zones with "Low" sheep density is proportionally lower than in those with high sheep density. There were no other significant differences. This has changed from the univariate analysis where there was a significant decreasing trend in the ratio of the number of inpatients from least deprived /number from most deprived areas, as sheep density decreases.

*Health board*: The ratio of the number of inpatients from least deprived to those from most deprived areas, in GR, HG and LO is significantly higher than in Greater Glasgow & Clyde (the reference health board). This means that in these health boards there are proportionally higher numbers of inpatients from least deprived areas than in Greater Glasgow & Clyde. The situation is inverse for AA, BR and LN. The health boards that were not significant in the univariate analysis (AA & BR) have now become significant with proportionally lower numbers of inpatients from least deprived areas than Greater Glasgow & Clyde. Tayside became not significant in the multivariate analysis, i.e. the ratio of the number of inpatients from least deprived /number of inpatients from most deprived areas, is not different from that in GC. These effects reflect to an extent the deprivation status of NHS board areas.

*Position* (*Latitude* (south to north) and *Longitude* (west to east)), *PWS's* and *Time of year* are no longer significant.

### 5.3.4.4 Risk factors associated with human campylobacteriosis hospitalisation in Scotland. Results from univariate and multivariate multinomial logistic regression

**(i) <u>Results from multinomial univariate logistic regression between hospitalisation classified by SIMD quintile</u>**

The interpretation of the following graphs is performed in the same way as those presented in Figure 4.11 for the analysis of reported cases (see Chapter 4, Section 4.3.3.4).

Figure 5.12 presents the results from the univariate multinomial logistic regression comparing *less* deprived (SIMD5, 4, 3 & 2) and *most* deprived (SIMD1) quintiles of hospitalisation for each risk factor.

*Gender*: There are proportionally lower numbers of female than male inpatients from SIMD4 and SIMD2 areas (Odds ratio's <1), whilst for SIMD3 and SIMD5 there are no significant differences (Figure 5.12(a)). Hence, the results appear to be internally inconsistent.

*Human population density*: The number of inpatients from less deprived (SIMD5, 4, 3 & 2) areas among the rural population is proportionally higher than that in the urban - reference - population (Figure 5.12(b)). Hence, there are proportionally higher numbers of inpatients from *less* deprived (SIMD5, 4, 3 & 2) areas among the rural than the urban population. The same occurs for the peri-urban population, except for the SIMD2 quintile that shows no significant difference.

*Longitude*: There are proportionally higher numbers of inpatients from less deprived areas in the East than in the West of Scotland (Figure 5.12(c)).

*Latitude*: There are proportionally higher numbers of inpatients from less deprived areas in the North than in the South of Scotland (Figure 5.12(d)).

*Age:* There are proportionally lower numbers of inpatients from less deprived areas (SIMD5, 4 & 2) in 25-64 years old than in 65+ years old, with no significant difference for SIMD3 (Figure 5.12(e)). There were no significant differences for the other comparisons.

*PWS's:* There are proportionally lower numbers of inpatients from *less* deprived (SIMD5, 4, 3 & 2) data zones where PWS's are not present than in data zones with PWS's (Figure 5.12(f)).

*Cattle density:* There are proportionally lower numbers of inpatients from *less* deprived (SIMD5, 4, 3 & 2) data zones with intermediate "Mid1" cattle density than from data zones with "High" cattle density (Figure 5.12(g)). The same result was expected for data zones with "Low" cattle density. However, this is only partly the case (SIMD 2,3 and 4 only) and hence the result is inconsistent.

*Sheep density:* There are proportionally lower numbers of inpatients from *less* deprived (SIMD5, 4, 3 & 2) data zones with "Low" and intermediate ("Mid1") sheep densities than in data zones with "High" sheep density (Figure 5.12(h)). The results are inconsistent for data zones with "Mid2" sheep density.

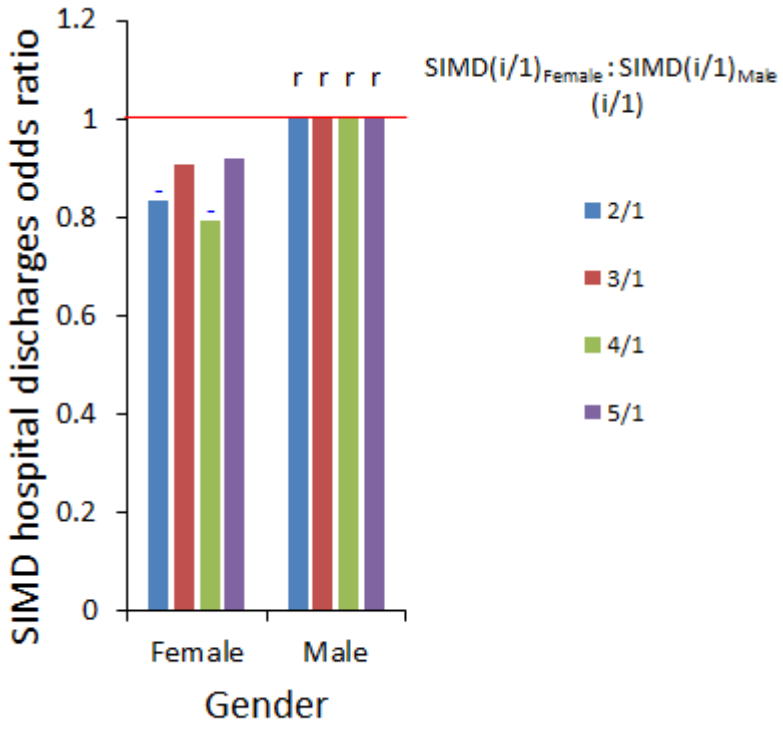
*Poultry density:* There are proportionally lower numbers of inpatients from less deprived data zones are also from areas with "Low", "Mid1" and "Mid2" poultry density than with "High" poultry density. However, this is not consistently statistically significant (Figure 5.12(i)).

*Time of year (season):* In the "summer" there are proportionally more inpatients from less deprived areas compared with the rest of the year. However, this is only statistically significant for the SIMD5/SIMD1 comparison (Figure 5.12(j)).

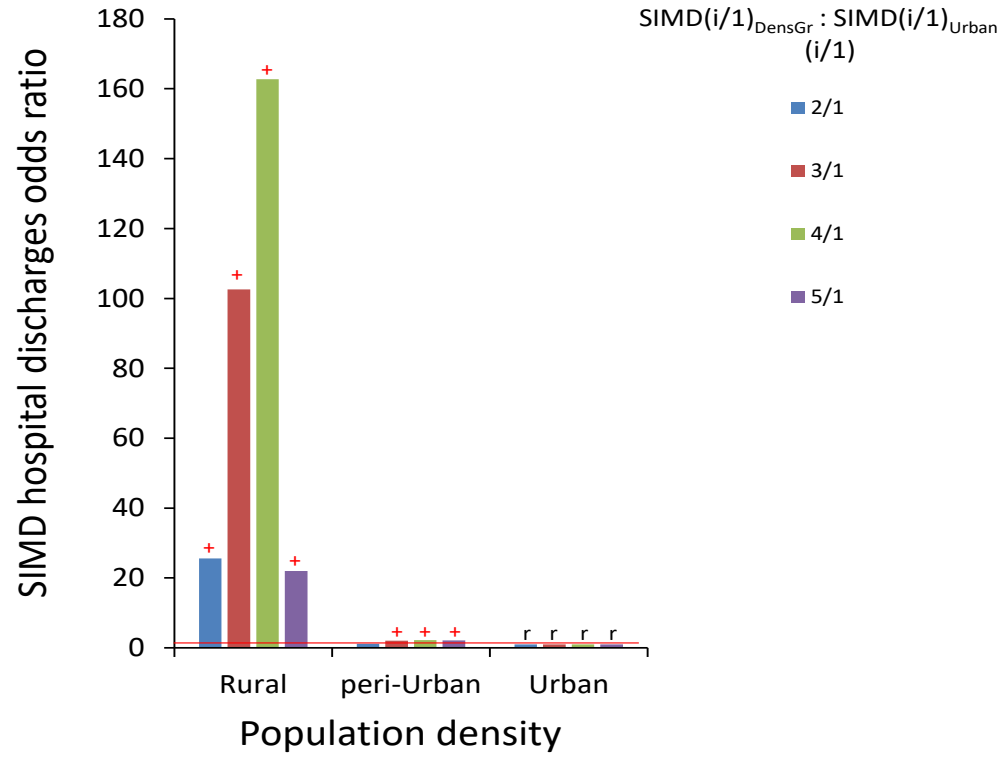
*Health board:* As for the binomial logistic regression, Greater Glasgow & Clyde health board was used as the reference in the multinomial logistic regression analysis (Figure 5.12(k)). Most of the comparisons (35/40) illustrate that there are proportionally higher numbers of inpatients from *less* deprived areas, compared with GC.

**Figure 5.12. Multinomial univariate logistic regression comparing hospitalisation in less deprived quintiles with the most deprived quintile, for each risk factor**

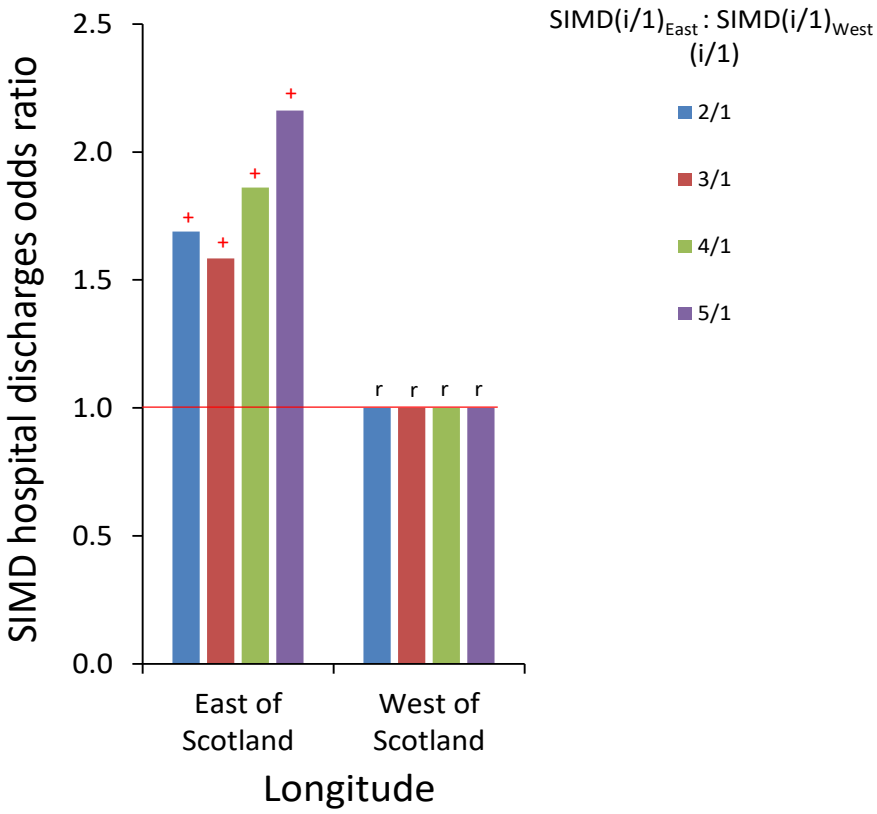
(a)



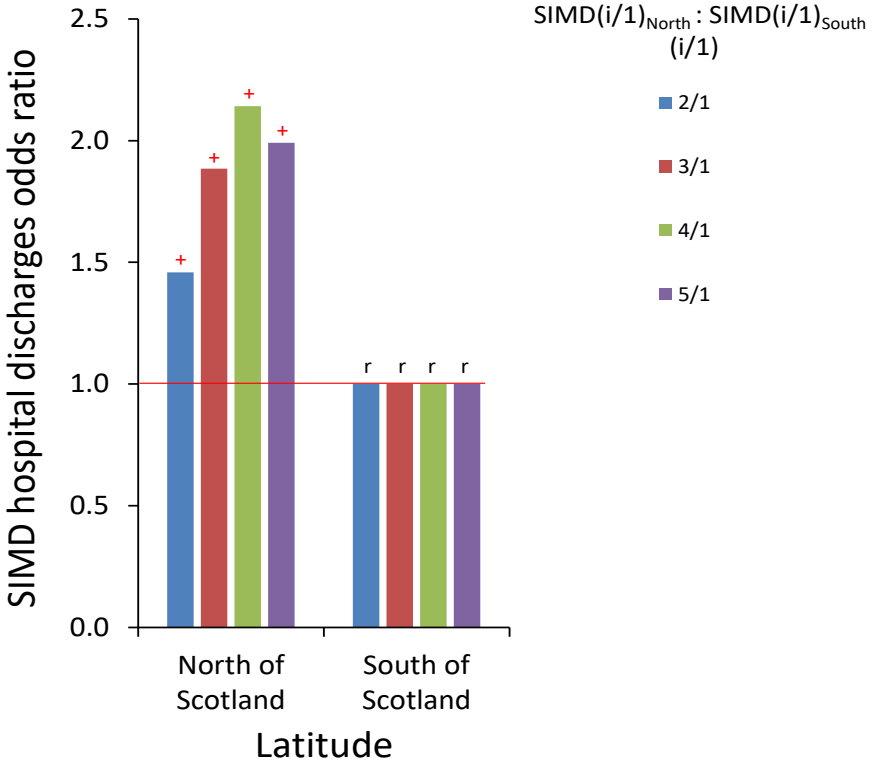

(b)



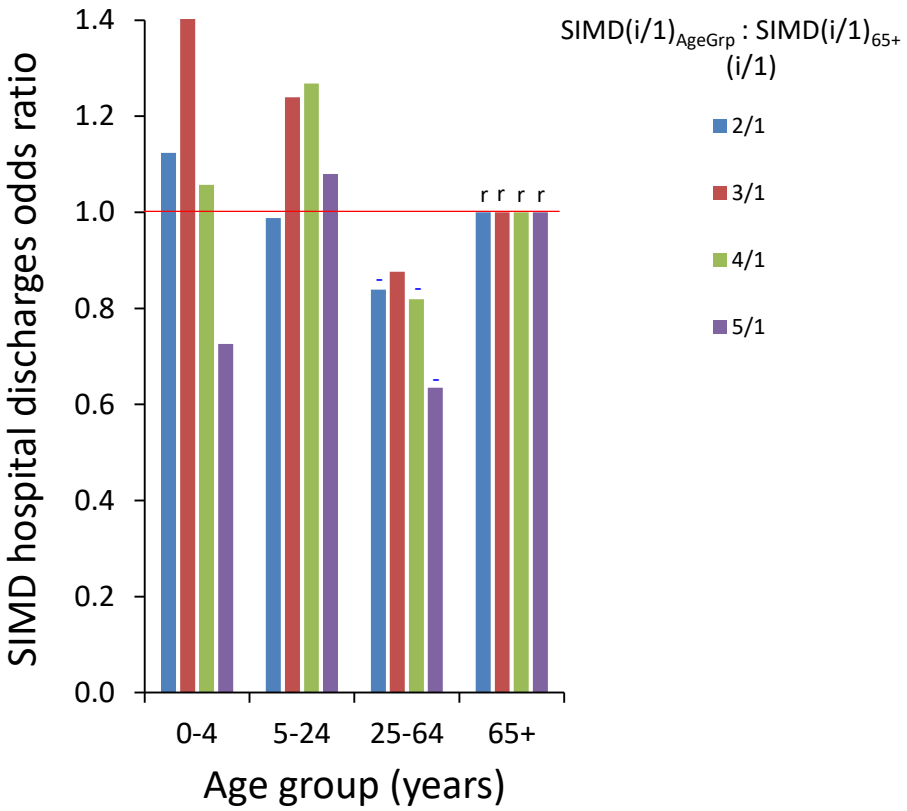
(c)


SIMD hospital discharges odds ratio
2.5
2.0
1.5
1.0
0.5
0.0
East of Scotland
West of Scotland
Longitude
SIMD(i/1)East : SIMD(i/1)West (i/1)
2/1
3/1
4/1
5/1

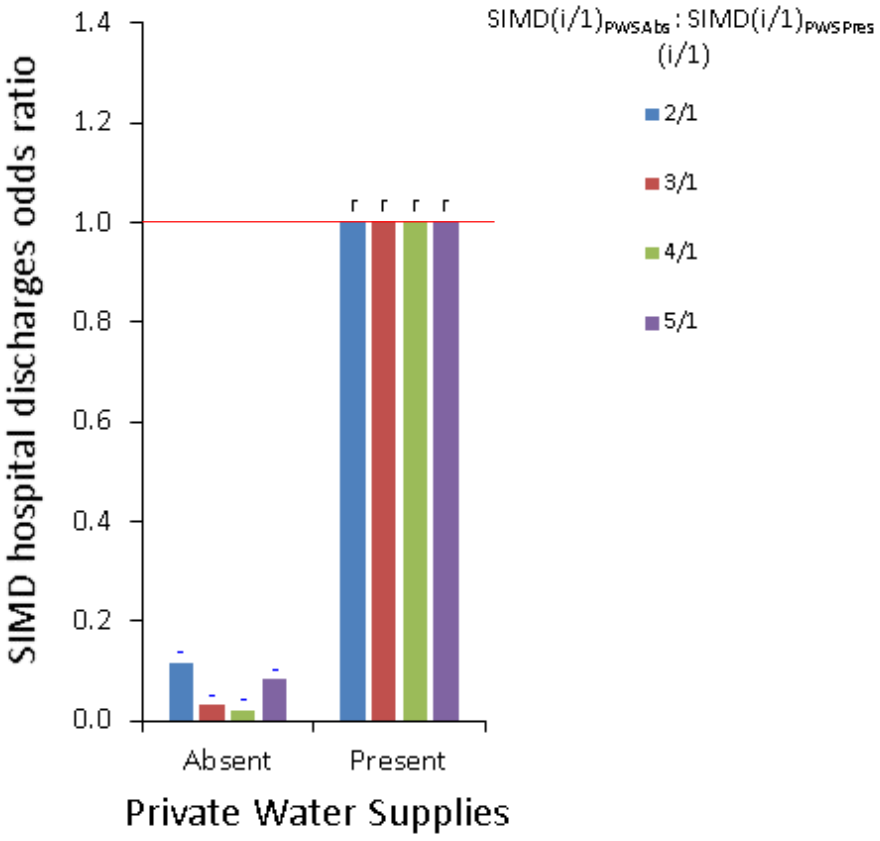
(d)


SIMD hospital discharges odds ratio
2.5
2.0
1.5
1.0
0.5
0.0
North of Scotland
South of Scotland
Latitude
SIMD(i/1)North : SIMD(i/1)South (i/1)
2/1
3/1
4/1
5/1

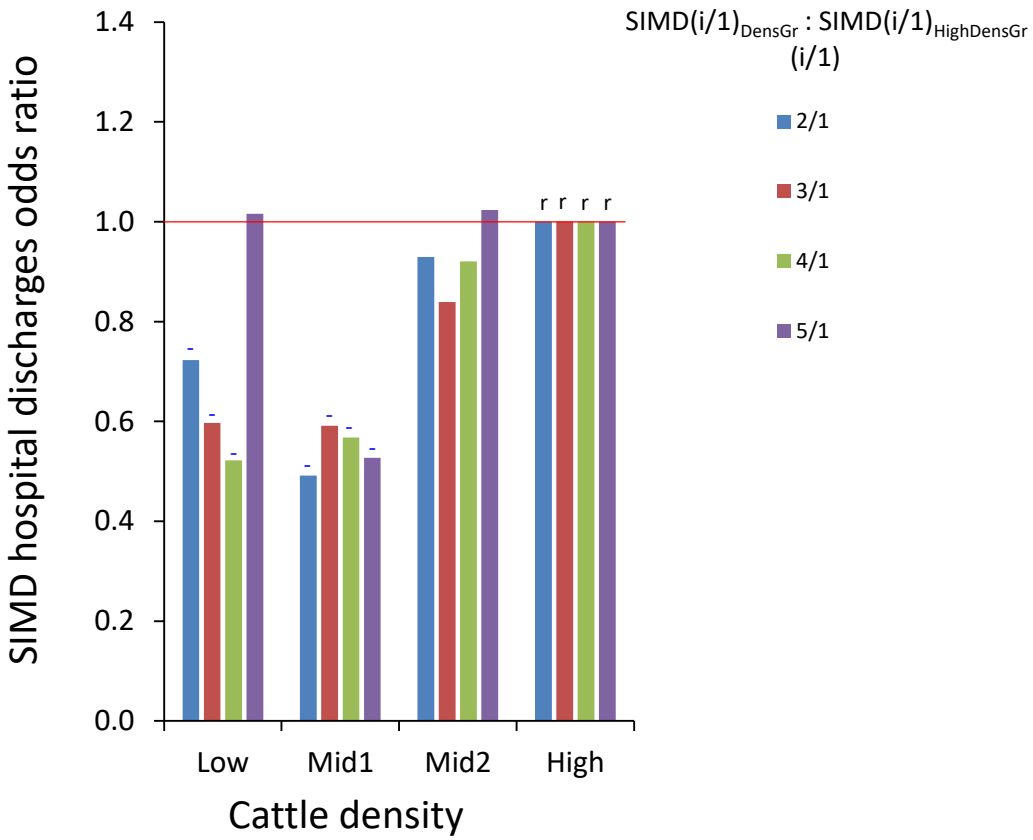
(e)


SIMD hospital discharges odds ratio
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
r r r r
0-4
5-24
25-64
65+
Age group (years)
SIMD(i/1)AgeGrp : SIMD(i/1)65+ (i/1)
2/1
3/1
4/1
5/1

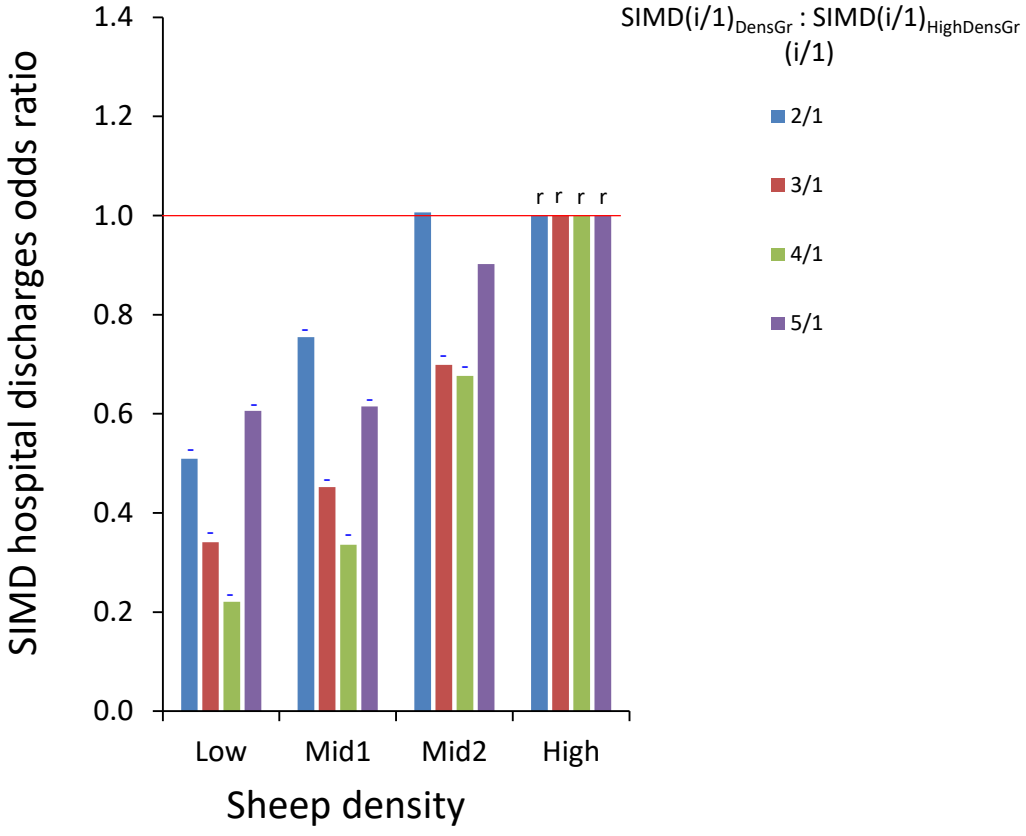
(f)


SIMD hospital discharges odds ratio
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
r r r r
Absent
Present
Private Water Supplies
SIMD(i/1)PWSAbs : SIMD(i/1)PWSPres (i/1)
2/1
3/1
4/1
5/1

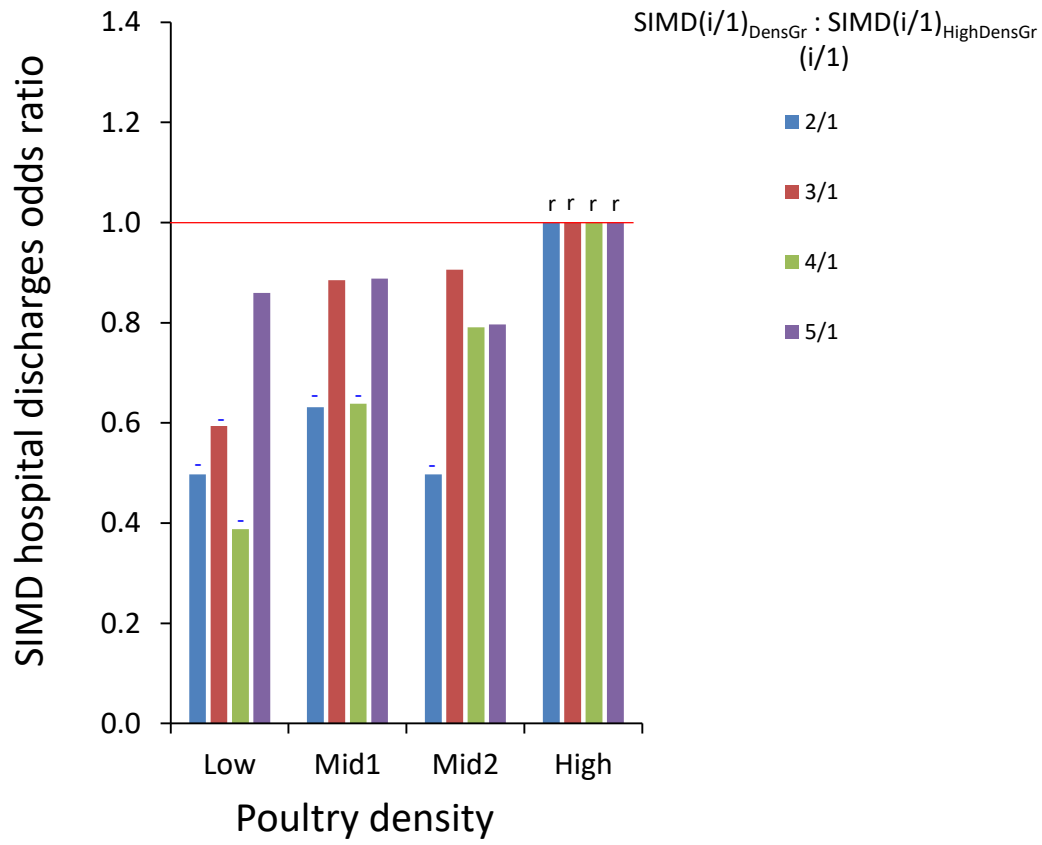
(g)


SIMD hospital discharges odds ratio
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
r r r r
Low
Mid1
Mid2
High
Cattle density
SIMD(i/1)DensGr : SIMD(i/1)HighDensGr (i/1)
2/1
3/1
4/1
5/1

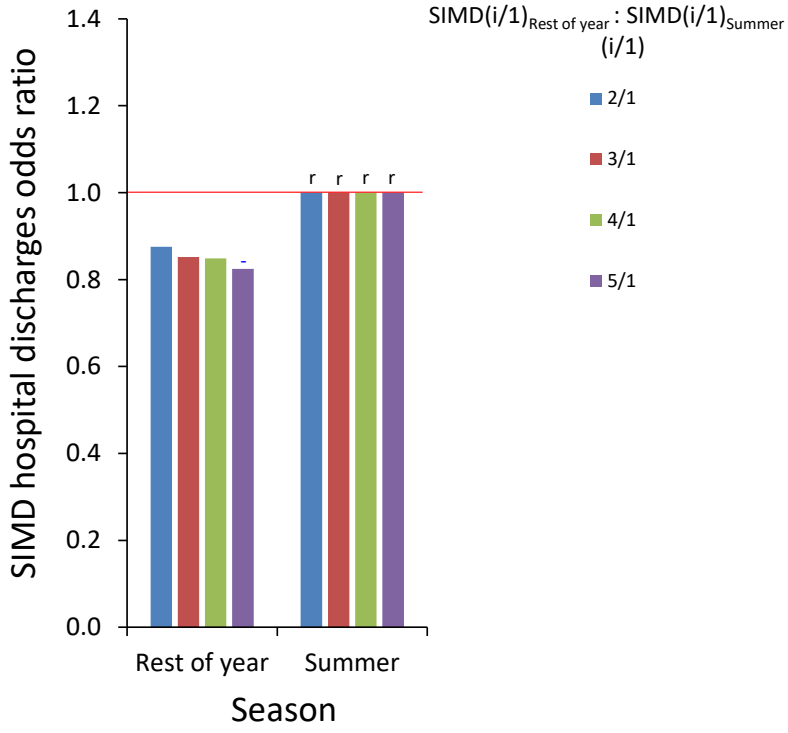
(h)


SIMD hospital discharges odds ratio
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
r r r r
Low
Mid1
Mid2
High
Sheep density
SIMD(i/1)DensGr : SIMD(i/1)HighDensGr (i/1)
2/1
3/1
4/1
5/1

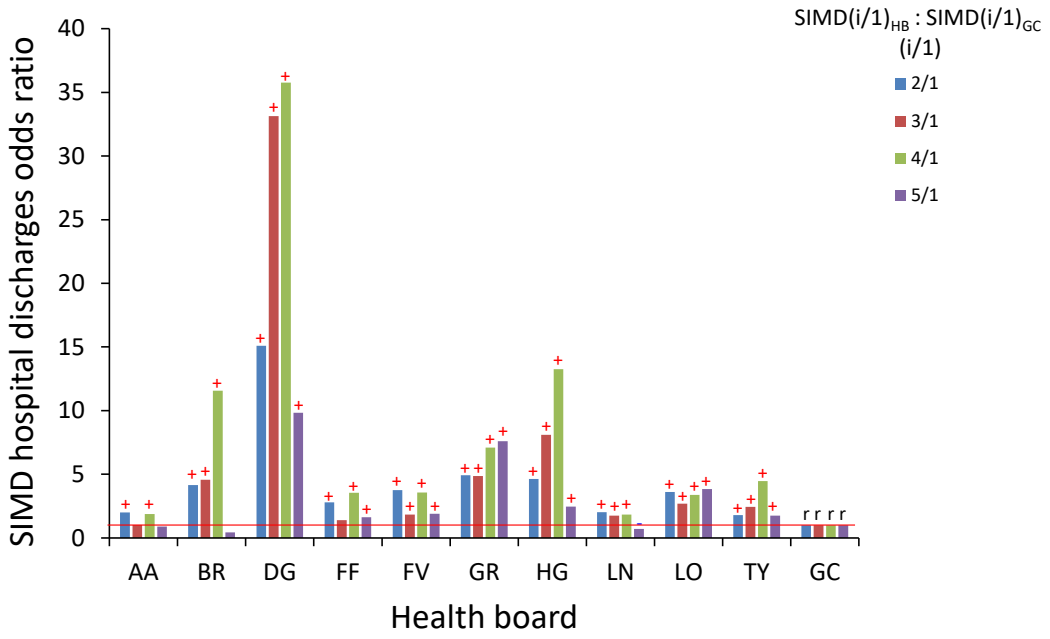
(i)


SIMD hospital discharges odds ratio
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
r r r r
Low
Mid1
Mid2
High
Poultry density
SIMD(i/1)DensGr : SIMD(i/1)HighDensGr (i/1)
2/1
3/1
4/1
5/1

(j)


SIMD hospital discharges odds ratio
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
r r r r
Rest of year
Summer
Season
SIMD(i/1)Rest of year : SIMD(i/1)Summer (i/1)
2/1
3/1
4/1
5/1

(k)




Multinomial univariate logistic regression comparing hospitalisation in *less* deprived (SIMD5, 4, 3 & 2) quintiles with the *most* deprived (SIMD1) quintile, for each risk factor: (a) gender, (b) population density, (c) longitude, (d) latitude, (e) age, (f) PWS's, (g) cattle density, (h) sheep density, (i) poultry density, (j) time of year (season) and (k) health board. Where SMID5 is least deprived and SIMD1 is most deprived. The letter "r" denotes the reference and "+" indicates the comparison is significantly higher whilst "-" indicates that it is significantly lower.

**(ii) Results from the multinomial multivariate logistic regression between hospitalisation classified by SIMD quintile**

These analyses were carried out and are presented in Annex 5.1.

The risk factors *gender* and *time of year* were removed during the analysis because they were not significant.

The following risk factors gave the same results as in the univariate analysis for all SIMD comparisons: *human population density*, *PWS's* and *health board* (HG vs. GC only).

For all the other risk factors, there were some differences from the univariate analysis and these are presented in Annex 5.1.

## 5.4 Analysis of spatial distribution of human campylobacteriosis hospitalisation relative to hospital geography

### 5.4.1 Aims

The aim of this section was to answer the following questions:

(iii) Are you more likely to go to hospital for campylobacteriosis if you live close to a hospital which reports cases of campylobacteriosis?
(iv) Does this depend on deprivation?

### 5.4.2 Data

Non-patient identifiable hospital discharge data from eDRIS were available for all patients discharged with a diagnosis of campylobacteriosis during the period 1st January 2012 to 31st March 2018 (see Chapter 5.2.1 & 5.2.2). The human population in each data zone was obtained from The Consumer Data Research Centre (https://data.cdrc.ac.uk/dataset/cdrc-2011-population-weighted-centroids-gb ). This also provided coordinates (easting and northing) of the centroid of each data zone. The SIMD quintiles for each data zone were obtained from http://www.gov.scot/Topics/Statistics/SIMD/DataAnalysis/Background-Data-2012.

The easting and northing as well as deprivation quintile for each hospital discharge and each member of the Scottish population was then allocated.

The names and addresses of each hospital (n=34, Apr 2017- Mar 2018) accepting and then discharging patients with campylobacteriosis in Scotland was obtained from ISD (http://www.isdscotland.org). The postcodes of the hospitals were geocoded (easting and northing) using the UK Grid Reference Finder (https://gridreferencefinder.com/).

### 5.4.3 Methods

The distance between the data zone of each person discharged and their closest hospital was determined. Then the distribution of all discharges within particular distances to their closest hospital was calculated (two distance intervals were used: 1km and 10km).

The "control" population comprised the same number of individuals as hospital discharges, but was randomly selected from the whole Scottish population. Their minimum distance from the closest hospital was calculated as above. As was the distribution of the number of controls within particular distances (1km or 10km intervals) to the closest hospital.

For the 1km distance interval, this "control" distribution was recalculated 500 times using the Monte Carlo method in PopTools (http://www.poptools.org/). From this, the average frequency distribution to the nearest hospital and 95% confidence intervals were calculated. This was repeated by SIMD quintile. If the confidence intervals did not overlap with the hospital discharge distribution then the results were considered to be significantly different.

The incidence of hospitalisation was also calculated at 10km intervals from the nearest hospital. This was repeated for each SIMD quintile.

### 5.4.4 Results and Discussion

Figure 5.13 shows a map of the 34 hospitals in Scotland which reported cases of campylobacteriosis during April 2017 - March 2018. Figure 5.14(a) shows that the distribution of the campylobacteriosis hospitalisation around hospitals generally follows the distribution of the "control" population. However there are some significant differences - e.g. at 1, 3, 4, 5 and 10 km there are more hospitalisations than expected. These differences account for 6.6% of total recorded campylobacteriosis hospitalisation during the study period.

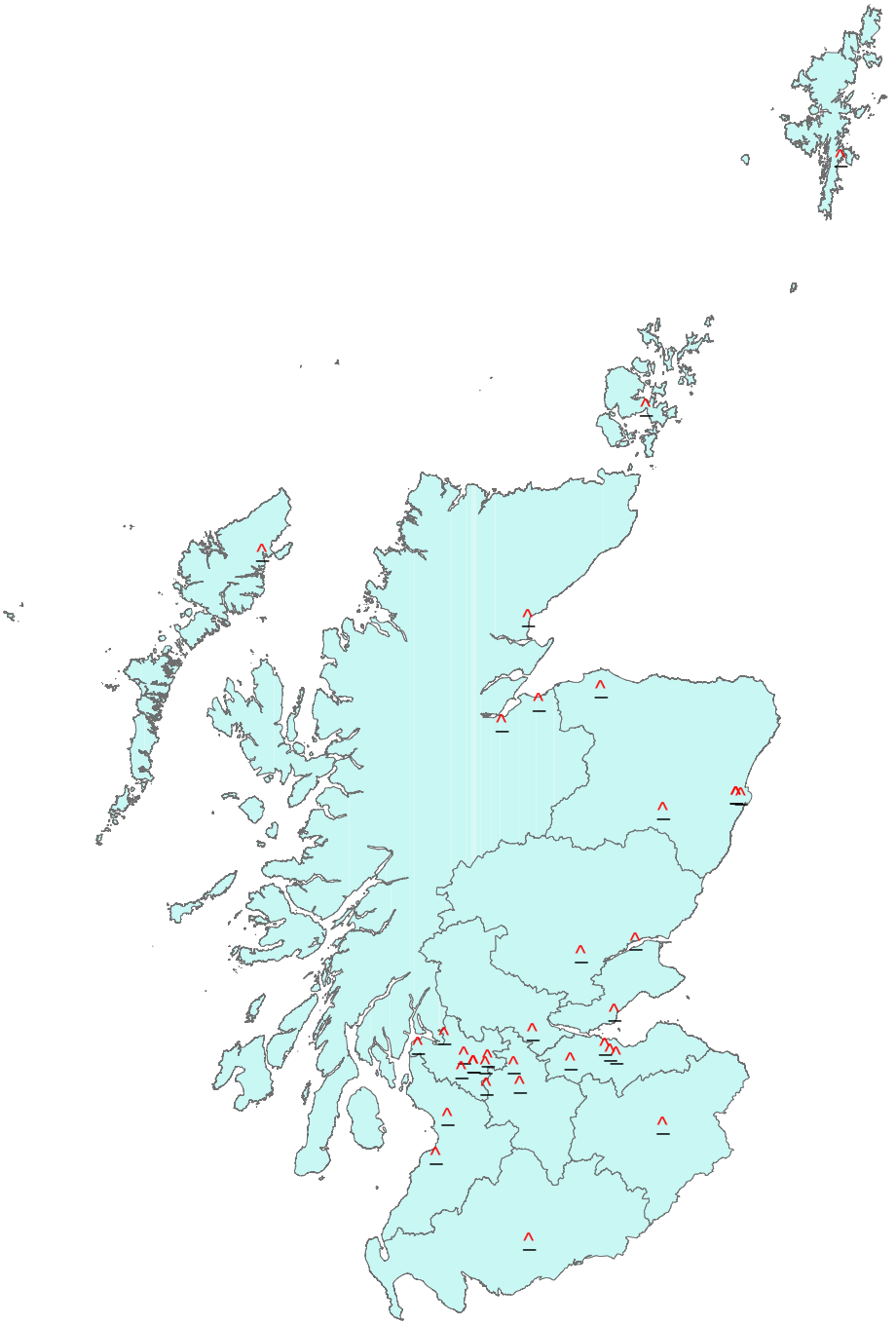
For all SIMD1 to SIMD5 the distribution of the campylobacteriosis hospitalisation generally follows the "control" population (Figure 5.14 (b) to (f)). Again, there are some significant differences in all graphs, within the first 10km, but there is no obvious difference by SIMD.

Figure 5.15(a) shows the rate of campylobacteriosis discharges as a function of residence distance from the hospital. The rate decreases with distance. There is a particular excess of discharges within the closest 10km of hospital (18.6 compared with 17.3 cases per 100,000) which corresponds to approximately 5.1% of all episodes. Figure 5.15 (b) to (e) also shows that there is an excess when considering each SIMD quintile. This ranges from 4.3 to 7.2%. However, this in itself would not account for the 9.2% excess in SIMD1 and 2 areas found in Figure 5.2 above.

That being said the *Campylobacter* discharge incidence is highest for SIMD1 and SIMD2 areas within 10 km of a hospital (Figure 5.15) and these areas have relatively high populations (Figure 5.16 (a)). Together, these factors account for the large number of SIMD1 and SIMD2 hospital discharges within 10km of a hospital (Figure 5.16 (b)).

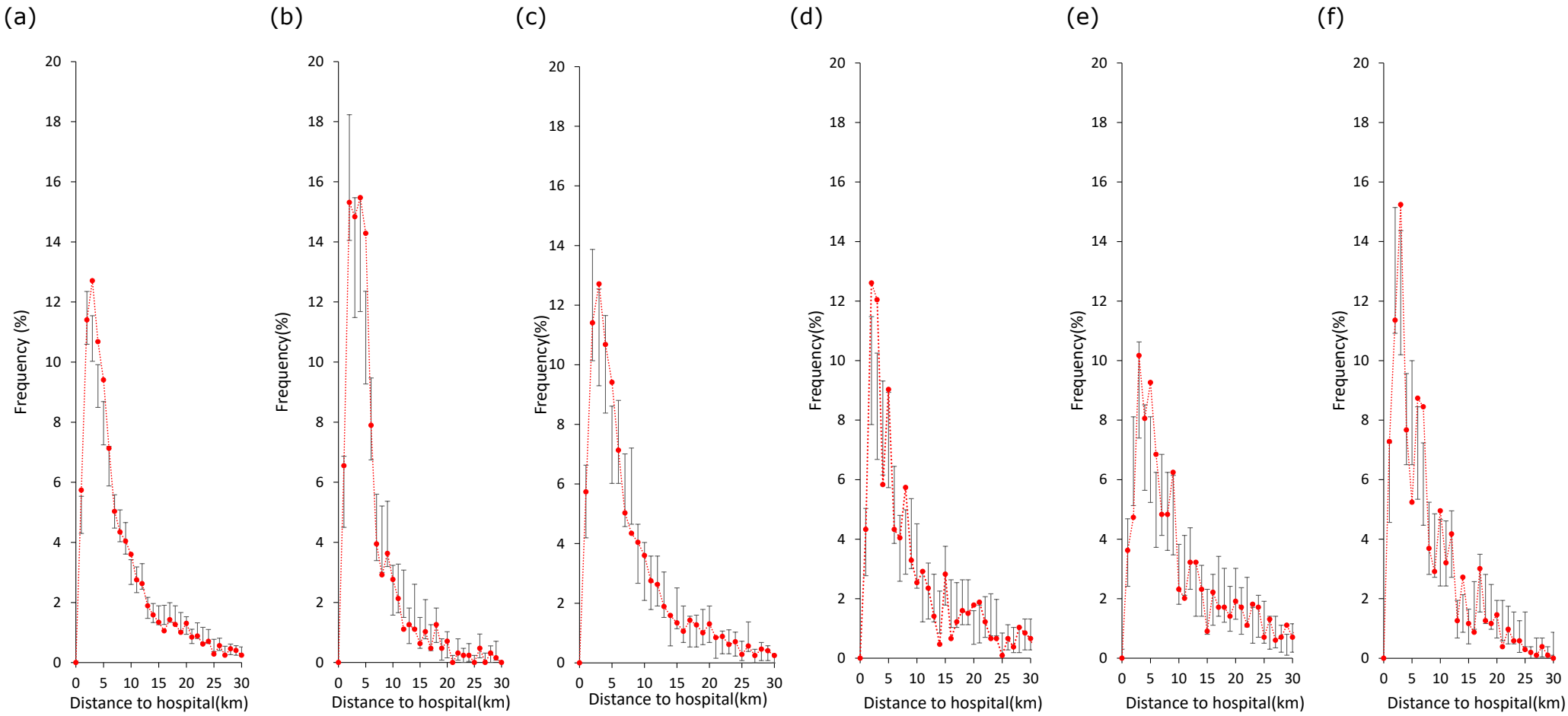
Hence, the high level of hospitalisation for residents of SIMD1 and SIMD2 areas appears to be due to the preponderance of them within 10km of a hospital combined with the high incidence rates there. It is worth noting that although SIMD5 has a relatively high population within <10km of a hospital, it has relatively low incidence of hospital discharges compared with SIMD1 and SIMD2.

**Figure 5.13. The geographical distribution of hospitals reporting campylobacteriosis cases in Scotland**



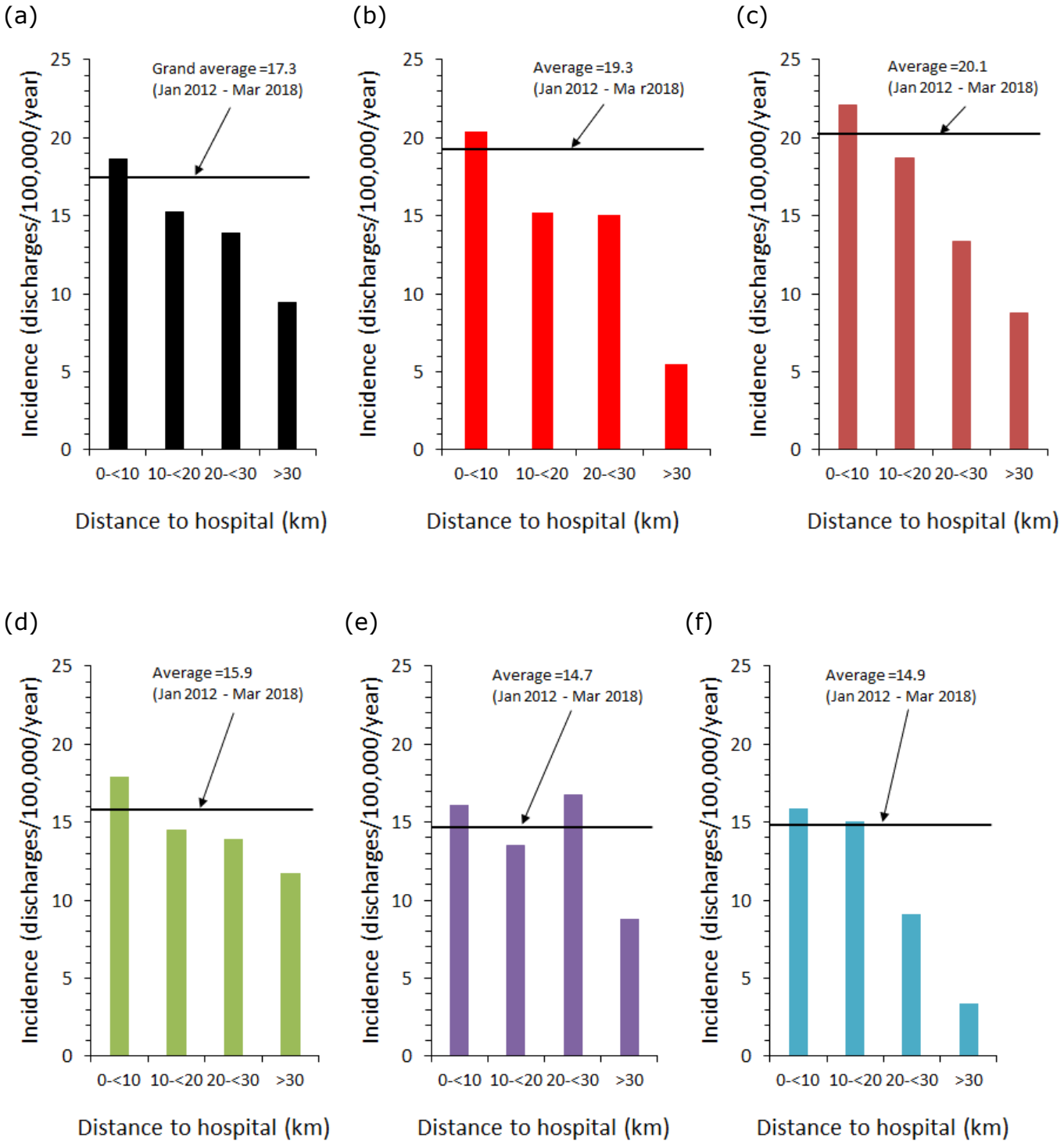
Mar 2018.

**Figure 5.14. Distribution of campylobacteriosis hospitalisation and "control population" relative to the distance to the closest hospital which reports campylobacteriosis cases**




(a) all hospitalisation vs. population; (b) SIMD1 hospitalisation vs. SIMD1 population; (c) SIMD2 hospitalisation vs. SIMD2 population; (d) SIMD3 hospitalisation vs. SIMD3 population; (e) SIMD4 hospitalisation vs. SIMD4 population and (f) SIMD5 hospitalisation vs. SIMD5 population. Hospitalisation are represented as (●) and confidence intervals are 95 percentiles for the control population (Note: for clarity in the graphs the average distribution of the "control population" is not represented by a symbol).

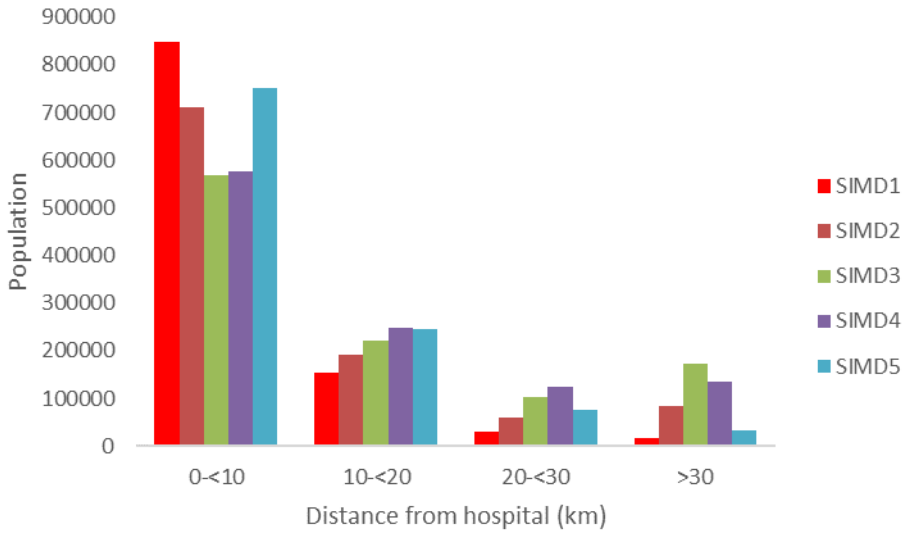
**Figure 5.15. Incidence of campylobacteriosis hospitalisation relative to the distance to the closest hospital which reports cases**



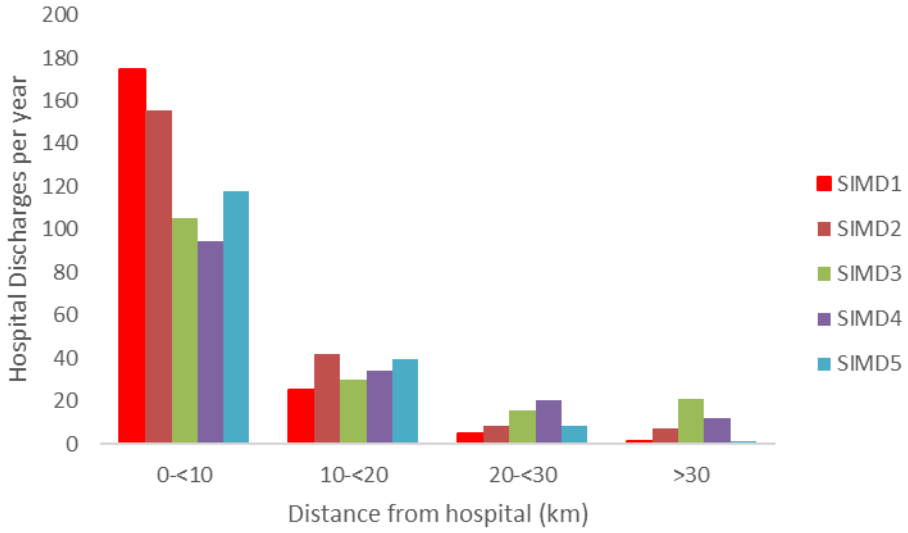

(a) all hospitalisation normalised by the total population at each distance. The following graphs are normalised by the relevant SIMD population at each distance: (b) SIMD1 hospital discharges; (c) SIMD2 hospital discharges; (d) SIMD3 hospital discharges; (e) SIMD4 hospitalisation and (f) SIMD5 hospital discharges.

**Figure 5.16 (a) Populations and (b) number of hospital discharges stratified by SIMD quintile and relative to the distance to the closest hospital which reports cases.**

(a)




(b)



## 5.5 Long term analysis of hospitalisation cases in Scotland

### 5.5.1 Aims

The aim of this section is to explore and understand the temporal dynamics of hospitalisation for campylobacteriosis in Scotland using time series analysis of data for 1990 to 2017. This is with a view to describing secular changes and offers an opportunity for generating hypotheses that may explain these changes.

### 5.5.2 Data

Summary data on hospitalisation with campylobacteriosis from Scotland during 1997 to 2011 (n=6,557) were available from the literature (N. J. C. Strachan, Rotariu et al. 2013). These data included information about the number of inpatient episodes stratified by five years age groups and the incidence rates for these groups (discharges/100,000/year). The numbers of inpatient episodes from 2012 to 2017 (n=5,646) were available from the current study (see Chapter 5.2.3.

Mid-year human population estimates (2012 to 2017) stratified by age and health board, were obtained from the National Records of Scotland (NRS) (https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates).

### 5.5.3 Methods

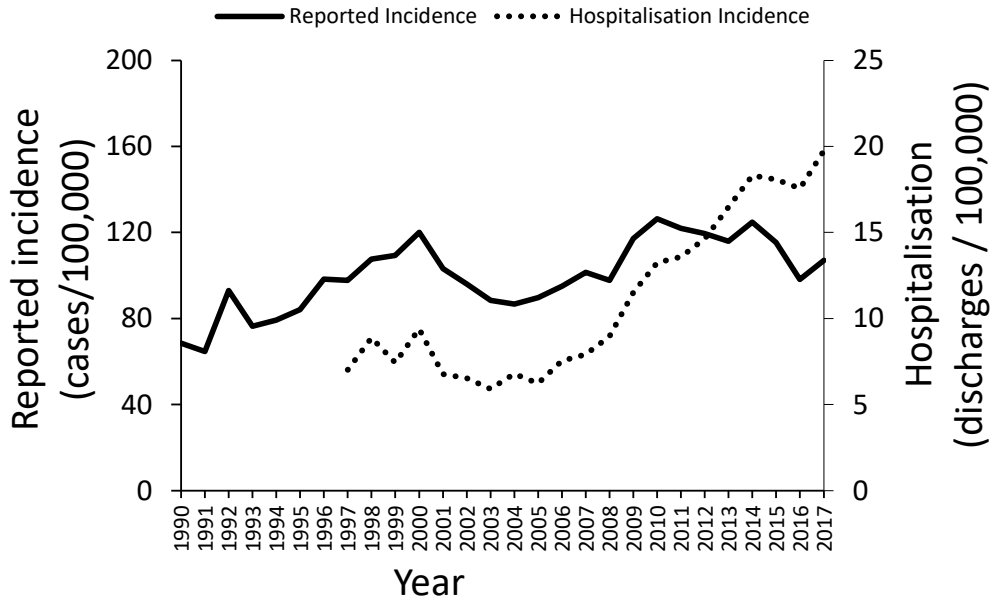
Campylobacteriosis hospital discharge data between 1997 and 2017 were plotted in terms of (i) overall incidence and (ii) incidence stratified by age group.

### 5.5.4 Results and Discussion

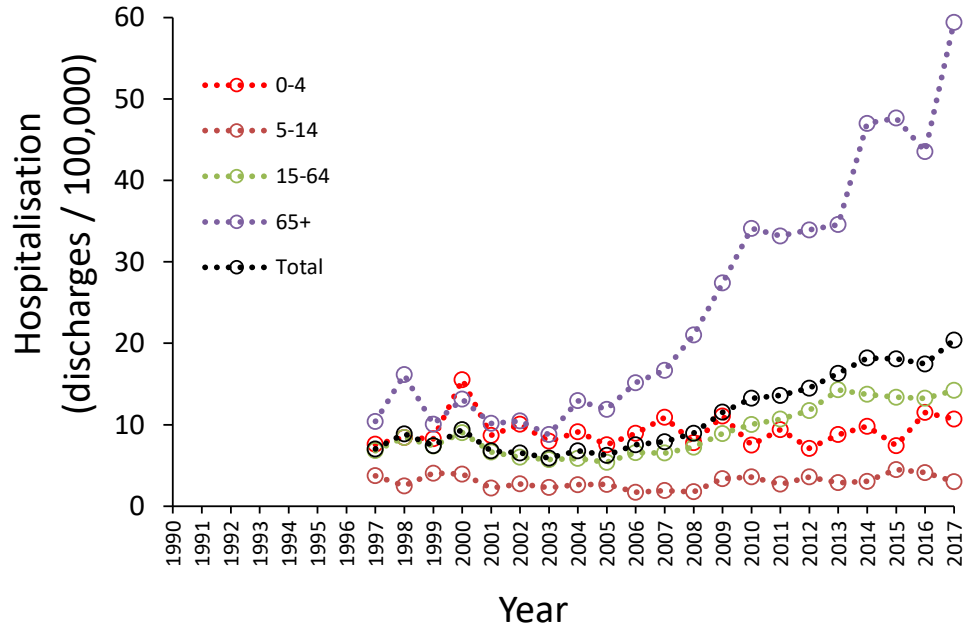
There has been a steady increase in the rate of hospitalisation with campylobacteriosis since 2005 (Figure 5.17(a)) resulting overall in a 3 fold increase. This has continued despite a decline in the reported incidence of campylobacteriosis during 2014 to 2016. The increase in the incidence of campylobacteriosis hospitalisation was pronounced in the 65+ year olds, with 0-4 year olds and 5-14 year olds being relatively stable (Figure 5.17(b)). In the adult population (15-64+ year olds) the increase in the rate occurred only between 2005 to 2013, the incidence rate being relatively stable afterwards.

**Figure 5.17. Long term analysis of campylobacteriosis hospital discharges**

(a)




(b)




(a) incidence of hospitalisation versus the incidence of reported cases and (b) incidence of hospitalisation stratified by age.

## 5.6 Overall discussion and conclusion

Descriptive epidemiology:

In total 15.5% of reported cases were hospitalised which is higher than previous studies for example, England and Wales approximately 10% (Gillespie, O'Brien et al. 2009), Scotland 7.1% (unpublished data from 2000-06), Spain 12.3% (Fajo-Pascual, Godoy et al. 2010) and New Zealand (1997 – 2008) 4.3% (Sears 2009).

In the current study there were 9.2% more hospitalisation admissions among patients from the two most deprived quintiles (SIMD 1 and 2) than the three *least* deprived. This is a change from 2000-2006 data where no difference was observed (unpublished). However, an excess of campylobacteriosis hospitalisation in the most deprived population has been reported previously in New Zealand (Sears 2009).

The rate of hospitalisation is higher among patients from urban and peri-urban areas (in total accounting for an excess of 14.5% of inpatient episodes). This excess hospitalisation could be due to one or more of several factors. For example: (i) proximity to a hospital; (ii) differing food preparation skills and/or consumption habits;(iii) different living conditions and/or diet resulting in lower immunity; (iv) different levels of co-morbidity and (v) acquired immunity may be more common in some rural areas.

With regard the first, from section 5.4 there is evidence that living closer to a hospital results in an excess of hospitalisation (6.6%). There are currently no supporting data for hypotheses (ii), (iii) and (iv). However, they may be worth further investigation. With regard hypothesis (v) there is evidence that immunity may be higher in rural areas (e.g. in the USA (Belongia, Chyou et al. 2003)).

Health boards vary in their rates of hospitalisation. For example, rates for BR, GR and HG are lower whilst GC is the highest. This is congruent with the above relationship with deprivation. (e.g. hospitalisation where 54% of GC's population live in the two most deprived data zones (SIMD1 & 2)).

On average, inpatients with *Campylobacter* stay in hospital for 3.7 nights. Fewer than 2% stay longer than 20 days. This is similar to the USA where there is a reported median stay of 3 days and fewer than 6% stay longer than 14 days (Scallan, Griffin et al. 2018). Similarly a median stay of 4 days was reported for patients in Finland with bacteraemia caused by to *C. jejuni* or *C. coli* (Feodoroff, Lauhio et al. 2011) and an average stay of 5 days was reported for patients in England and Wales (Gillespie, O'Brien et al. 2002).

Poisson regression models:

Both univariate and multivariate Poisson regression models show that increasing deprivation is associated with increasing rate of admission in campylobacteriosis. This agrees with the descriptive analysis carried out above (Figure 5.2) and with other studies on hospital admission rates for gastrointestinal infections in the UK (Olowokure, Hawker et al. 1999) as well as campylobacteriosis hospitalisation in New Zealand (Sears 2009).

Increasing human population density was also associated with increasing incidence of hospitalisation in the univariate analysis, whilst it was opposite in the multivariate analysis. However SIMD score is correlated with population density i.e. higher proportion of deprived people in areas of high population density (data not presented).

Higher incidence of hospitalisation was associated with the west in the univariate regression model, but not in the multivariate model. Also, higher incidence of hospitalisation was associated with the south in both univariate and multivariate regression models. Interpretation of the map in Figure 5.9 points to considerable heterogeneity. This is likely due to the low numbers in each data zone (on average less than one hospitalisation per data zone). A methodology for combining data zones into larger areas has not yet been developed. This would potentially be a more fruitful first step to interrogating these data.

It is unclear why increasing cattle, sheep and poultry densities all appear to be protective in both univariate and multivariate models as these animals are all known to be *Campylobacter* reservoirs in Scotland (Ogden, Dallas et al. 2009). Several potential hypotheses could provide an explanation.

First, the strains from these sources may be less pathogenic and thus less likely to cause infection sufficiently severe to require admission. However, it is known that a number of the *Campylobacter* sequence types found in these animal populations are also found in cases of human disease (Sheppard, Dallas et al. 2009).

Second, cases caused by these sources tend to be in rural areas not close to a hospital. As mentioned above, there is some evidence to show that hospitalisation rates are higher closer to a hospital. However, since the hospitalisation excess is only 6.6% this is unlikely to be the whole explanation.

Third, infections from these sources contribute only a fraction of campylobacteriosis cases – the most important vehicle being food such as chicken (Wagenaar, French et al. 2013). This could explain why they would not be risk factors but would not account for them being protective.

Fourth, the population in areas with higher densities of these animals have acquired immunity. Previous work has indicated that people with occupational animal exposure are less likely to become ill when exposed to *Campylobacter* (Forbes *et al* 2009) and that there are higher levels of seropositivity in farm residents (Belongia, Chyou et al. 2003). One might also expect a higher incidence of the disease in young children, when they are first challenged with this pathogen, but that incidence reduces with age as immunity is acquired (Havelaar, van Pelt et al. 2009). The Poisson regression does not include an analysis by age but it is known that in rural areas in Scotland young children have a higher incidence of infection (N. J. Strachan, Gormley et al. 2009).

Similar arguments to the above can apply to explain why private water supply density is protective. It is known that private water supplies tend to be in rural areas where there are farm animals (Smith-Palmer, Cowden 2010) and contamination can occur from the faeces of these animals (and also possibly wild birds and other wildlife that may be present). For example in a north-east

Scotland study 62% of PWS were contaminated with coliforms compared with 1.7% of mains supplied water samples (Smith-Palmer, Cowden 2010).

Logistic and Multinomial Regression:

*Logistic*

The logistic regression points to risk factors that discern between the least and most deprived populations. There is variation in all of the risk factors between univariate and multivariate regression except for two. The first is population density, where there are proportionally more inpatients from least deprived data zones in rural and peri-urban areas compared with urban ones. This could be due to a higher number of deprived people being present in urban areas. The second is cattle density which does not show a trend with deprivation and is therefore inconclusive.

In areas with high sheep density both models found that there are proportionally, more inpatients from least deprived data zones than in areas of low sheep density. This again reflects the situation that the most deprived population tends to be resident in urban areas where there are likely to be fewer sheep. However, it is surprising that you do not see this pattern for cattle although for the general public access to sheep is generally greater as in parts of Scotland they are both free to roam over wide areas.

In both the univariate and multivariate analysis the GR, HG and LO health boards have proportionally more inpatients from least deprived areas compared with the reference Greater Glasgow & Clyde (highest hospital episode incidence rate). This may be explained by a larger proportion of their populations living in the least deprived compared with the most deprived SIMD quintiles (Table 2.2). Consistency of findings was also observed in the uni- and multivariate analysis for LN but in the opposite direction.

*Multinomial*

The following risk factors gave the same results for both univariate and multivariate multinomial regression: *human population density*, *PWS's* and *health board* (HB). They showed proportionally more inpatients from *less* deprived data zones, the more rural the area, where PWS's were present and in Highland compared to the reference health board GC. These results are not surprising because of the distribution of the population (i.e. most deprived in urban areas and where there are no PWS's. Also GC has a higher proportion of patients in most deprived and fewer in least deprived areas than HG (Figure 2.2)).

Proximity of inpatients to hospitals:

There was an excess of 5.1% within 10 km of the nearest hospital. This is in agreement of a number of reports that hospitalisation rates increase the closer the patient's home is to the hospital. For example in a GIS analysis of all hospitalisations across three health regions in British Columbia, Canada (Lin, Allan et al. 2002) and in cardiac vascular services in New Jersey, the USA (Gregory, Malka et al. 2000). There are some studies however which do not show this effect. For example in Denmark, inpatient hospital admissions was not associated with distance to the hospital (Bech, Lauridsen 2009). It is likely that

the distance relationship may be sensitive to the population structure, severity of the symptoms as well as the health seeking behaviours of the population.

The excess of hospitalisation within 10km was also observed for all of the deprivation quintiles. However, the reason for the high rate of hospitalisation within SIMD1 and SIMD2 (9.2% excess reported earlier) appears to be due to the large populations within <10km of a hospital combined with the high incidence rates within those populations. It is unclear why the incidence should be higher for SIMD1 and SIMD2. However, it is known that these populations generally have poorer health and it can therefore be hypothesised that this in combination with a *Campylobacter* infection may lead to a higher hospitalisation rate. Further work on investigating the general health of these hospitalised cases would enable testing of this hypothesis.

It is assumed that the hospital that the person attends in the current study is the closest to their home data zone. It is likely that this is not always the case but data were not available to identify whether this would have a significant effect on the results obtained.

Long term variation in hospitalisation:

The incidence of hospitalisation with *Campylobacter* infection has increased between 1997 and 2017 by 181%. This increase has been greatest in those older than 65 years (472%). It is unclear why there has been this dramatic increase in the elderly but it is known that the use of PPIs in this group has increased substantially during this period (N. J. C. Strachan, Rotariu et al. 2013). This may be an indicator of poor gut health and/or may have increased the opportunity for human campylobacteriosis infections through increased stomach pH. There is also the possibility that there has been increased exposure to *Campylobacter* as consumption of chicken has increased during this period (N. J. C. Strachan, Rotariu et al. 2013) but one might then expect this to be observed across all of the age groups.

Conclusions

The incidence of campylobacteriosis hospitalisation has increased three-fold since 2005. This is mostly due to the increase affecting the elderly (>65 years), whilst hospitalisation rates for children (both <5 years and the 5-14 year age groups) have been relatively stable throughout.

Focussing on deprivation, there is a 9.2% excess of hospitalisation with campylobacteriosis among residents of the *most* deprived (first two) SIMD quintile areas. Deprivation is positively associated with hospitalisation with campylobacteriosis.

The reason for the high level of hospitalisation within SIMD1 and SIMD2 (9.2%) appears to be because of two factors. First, the large SIMD1 and SIMD2 populations within <10km of a hospital and second the high incidence rates within those populations. It is unclear why the incidence rate is so high but it may be that other health conditions within these populations are a contributing factor.

Comparing SIMD1 and SIMD5, shows that proportionally there are more patients from rural and peri-urban areas hospitalised with campylobacteriosis compared

with those from urban populations. This may be due to proportionally higher numbers of least deprived individuals living in rural and peri-urban compared with urban areas. In areas with high sheep density there are proportionally more patients from least deprived areas hospitalised with campylobacteriosis.

There are also proportionally more patients from *less* deprived areas hospitalised with campylobacteriosis in more rural areas, where PWS were present and in Highland compared to the reference health board, GC.

There is evidence to suggest that living close to a hospital increases the likelihood of being hospitalised. This does not vary by deprivation.

# 6. The Case and Control Questionnaire Datasets

## 6.1 Introduction

This chapter describes the datasets used in the case-control study (Chapter 7) and the case-case study (Chapter 8). It details how the studies were carried out, the definitions of cases and controls, the participation of the health boards, how the data were processed, response rates, quality assurance and any biases that occurred.

## 6.2 Questionnaires, covering letter and information leaflets for NHS boards

Packs containing questionnaires, covering letter and leaflets for both cases and controls (see Section 2.2 and Annexes 2.1 and 2.2) were delivered to each of the participating health boards (Table 6.1) when they were ready to start the study. Additional packs were sent as and when required.

## 6.3 Submission by NHS boards of case and control paper questionnaires

### 6.3.1 Procedure of case and control selection

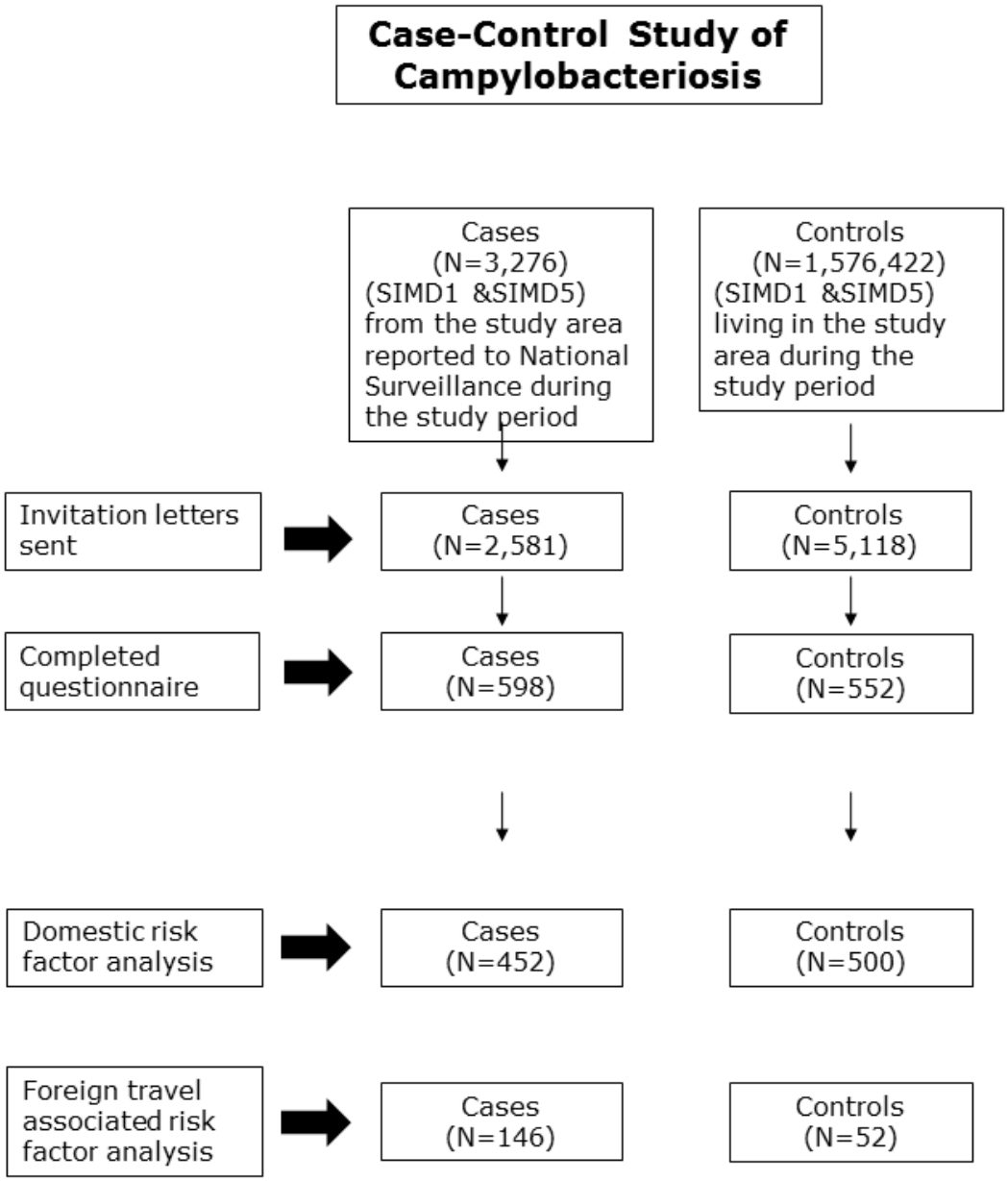
**Cases** were defined as any person above five years of age, living in the study area, not part of a known outbreak, with a culture-confirmed *Campylobacter* infection. Cases were identified by the Health Protection Team (HPT) of the NHS board of residence. SIMD for each case was identified from the postcode. Only those cases in 1$^{st}$ (most deprived) and 5$^{th}$ (least deprived) SIMD quintiles were selected. Potential participants were approached by means of a standard invitation letter bearing the NHS letterhead and signed by the local Consultant in Public Health Medicine (CPHM) , an information sheet, a consent form and questionnaire were also attached (Annex 2.1 and 2.2). On the questionnaire an ID number was included. The participant could complete the questionnaire and return it by post to NHS Tayside or alternatively, using the ID number, log on to a University of Aberdeen secure website and complete the consent form and the questionnaire online.

**Controls** were defined as any person above five years of age that had not had diarrhoea and or vomiting in the previous seven days, living in the study area and in one of the least or most deprived SIMD data zone quintiles. Controls were randomly selected by the participating Health Protection Teams using the Community Health Index (CHI) number (this is a unique ten digit number used by the National Health Service when registering all patients in NHS Scotland). They were selected following the same weekly distribution as observed for the reported cases by each health board in order to account for seasonal patterns.

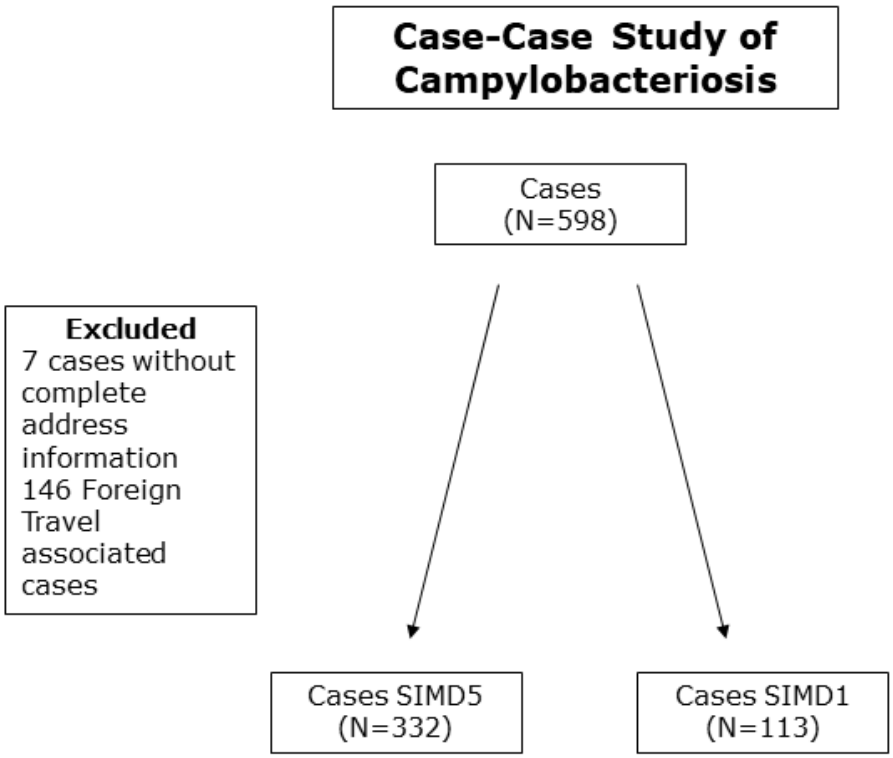
Potential control participants, two were selected for each case, were approached in the same way to that used for case participants by being sent an invitation letter, information sheet, a consent form and questionnaire (Annex 2). The controls could respond in the same way as cases either by post or through the secure University of Aberdeen website.

**Figure 6.1 (a) Case-control and (b) case-case study flow charts.**

(a)



(b)




---

Fig. 6.1(a) shows a flow chart of the case-control study. There were 3,276 campylobacteriosis cases >5 years old, from SIMD1 and SIMD5 quintiles from the study area during the period (Table 6.1). Of these, questionnaires were sent to 2,581 cases and completed responses were obtained from 598. This comprised 452 domestic cases and 146 that had travelled abroad out with the United Kingdom.

The control population (N= 1,576,422) comprised all individuals resident in the study area and aged over 5 years. During the study, control questionnaires were sent to 5,118 individuals of which completed responses were returned from 552. Of these 500 were considered to be domestic cases whilst 52 reported foreign travel.

Two sets of case-control analysis were performed, the first for domestic cases and the second for foreign travel associated cases.

The case-case study is depicted in Fig 6.1(b). This used the cases from the case-control study where address information was available and with the foreign travel associated cases removed. This was in order to identify what the differences were within the domestically acquired cases. This left 445 cases which were stratified into SIMD5 (n=332) and SIMD1 (n=113) cases.

### 6.3.2 Participation of NHS boards

Each health board submitted questionnaires to 2 controls for every single case.

Eight health boards (Table 6.1) sent out questionnaires (FF, FV, GR, GC, HG, LO and TY). Lanarkshire sent out questionnaires at the start of the project but withdrew from the study as explained in Chapter 2. Electronic versions of the questionnaires were also available online for cases and controls to complete. Table 6.1 shows the duration that each health board participated in the case-control study.

**Table 6.1 Starting dates and participation (in months) for the case-control questionnaire study by health board.**

| Health board | Case - control study starting date | Number of months participating Q'aires submitted |
|---|---|---|
| **Fife** | 1/7/16 | 26.00 |
| **Forth Valley** | 7/3/17 | 17.75 |
| **Glasgow & Clyde** | 1/6/16 | 27.00 |
| **Grampian** | 1/6/16 | 27.00 |
| **Highland** | 9/7/16 | 25.7 |
| **Lanarkshire** | 15/8/16 | -* |
| **Lothian** | 15/8/16 | 24.50 |
| **Tayside** | 15/6/17 | 14.50 |

*Lanarkshire started sending questionnaires but withdrew from the project.

## 6.4 Data entry and processing

Completed postal questionnaires (545 case and 505 control) returned to NHS Tayside were anonymised and uploaded into SNAP (Survey aNAlysis Package, electronic survey database). These were then combined with the questionnaires (53 case and 47 control) that were completed online by participants and then all were uploaded to DaSH (NHS Grampian/University of Aberdeen safe haven). Questionnaires were checked for completeness by ensuring that 14 questions for cases and 12 from controls that were mandatory were complete. Responses to each question in each questionnaire were then checked for validity (i.e. respondents had answered the question posed).

Just under 9% of the questionnaires were submitted electronically. This was lower than anticipated - perhaps because the questionnaire and reply paid envelope were provided (thus completing online meant disposal of the paper questionnaire and reply paid envelope).

## 6.5 Return rate and quality assurance

### 6.5.1 Return Rates

Table 6.2 presents the number of questionnaires submitted and the return rates for each health board. Overall 7,699 questionnaires were sent to both cases and controls. The return rate was 22.7% for cases and 10.6% for controls.

The return rates by deprivation in the current study were: (i) 14.0% for SIMD1 cases and 5.4% for SIMD1 controls, respectively (see Table 6.3) and (ii) 27.7% for SIMD5 cases and 15.7% for SIMD5 controls, respectively (see Table 6.4).

There were < 5 questionnaires returned from Lanarkshire. These were retained in the study.

Response rates from cases and controls were lower than expected. Questionnaires from a previous study sent by post to campylobacteriosis cases from Grampian achieved a response rate of 34% (Anon. 2017) and this may be due to the length of the questionnaire (12 pages compared with 4 pages). A campylobacteriosis case-control study, involving a postal questionnaire conducted in Aberdeenshire and Moray during 2005 to 2007 achieved 59.1% and 37.0% response rates for cases and controls respectively (Smith-Palmer, Cowden 2010). A recent case-control study in Denmark among children and young adults achieved response rates of 58% and 61% for cases and controls respectively (Kuhn, Nielsen et al. 2018). This Danish study recruited participants by post and completion of questionnaire online. Two postal reminders were sent 7 and 14 days after the initial invitation if required. It is unclear why the current study had lower response rates than anticipated. It may be that the current Scottish population are more frequently asked to respond to surveys and so there is a reticence to complete (Moy, Murphy 2016). However, in Denmark the response rates were very high and it may be that the double reminder may have been helpful. Reminder letters were considered for the current study but previous experience had shown that these were unlikely to have a great impact on the response rate.

**Table 6.2 Case-control questionnaire submission numbers and return rates by participating health board.**

| | Returned/Submitted | | Returned/Submitted | |
|---|---|---|---|---|
| Health board | cases | (%) | controls | (%) |
| Fife | 39/179 | 21.8% | 32/358 | 8.9% |
| Forth Valley | 20/75 | 26.7% | 8/150 | 5.3% |
| Greater Glasgow & Clyde | 151/976 | 15.5% | 164/1986 | 8.3% |
| Grampian | 108/432 | 24.8% | 144/868 | 16.6% |
| Highland | 25/86 | 29.1% | 24/170 | 14.1% |
| Lothian | 168/584 | 28.8% | 120/1129 | 10.6% |
| Tayside | 75/249 | 29.7% | 52/457 | 11.4% |
| Total | 586[a]/2581 | 22.7% | 544[b]/5118 | 10.6% |

[a] The total number of cases in the study was 598. There were 12 cases either from Lanarkshire or of unknown health board which are not included in the above table. It is not possible to provide exact numbers because the cases would encompass numbers <5.
[b] The total number of controls in the study was 552. There were 8 cases either from Lanarkshire or of unknown health board which are not included in the above table. It is not possible to provide exact numbers because the cases would encompass numbers <5.

**Table 6.3 Case-control SIMD1 and SIMD 5 questionnaire submission numbers and return rates by participating health board.**

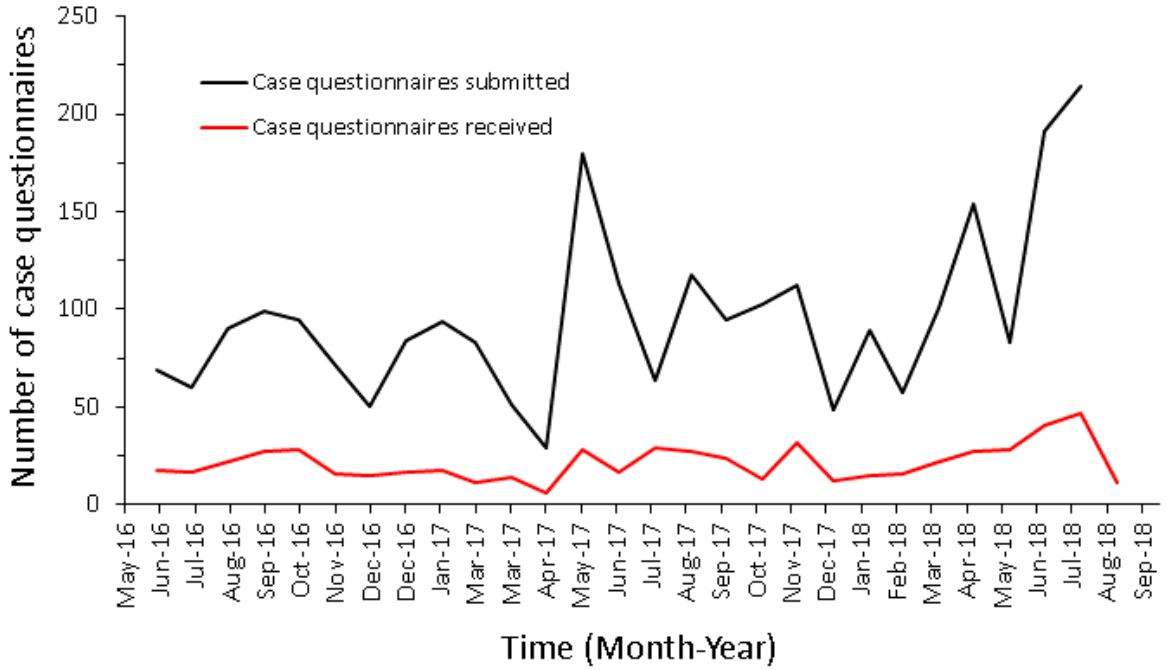
| | Returned/Submitted | | Returned/Submitted | |
|---|---|---|---|---|
| SIMD Quintile | cases | (%) | controls | (%) |
| SIMD1 | 135/952 | 14.0% | 139/2544[b] | 5.5%[a] |
| SIMD5 | 455/1629 | 27.7% | 407/2574[b] | 15.8%[a] |
| Grand total | 590[a]/2581 | 22.9% | 546[b]/5118 | 10.7% |

For confidentiality where there are less than 5 individuals in a category they are denoted as <5.
[a] The total number of cases is 598 but since SIMD information is missing in 8 the total in the table is 590.
[b] The total number of controls is 552 but since SIMD information is missing in 6 the total in the above table is 546.

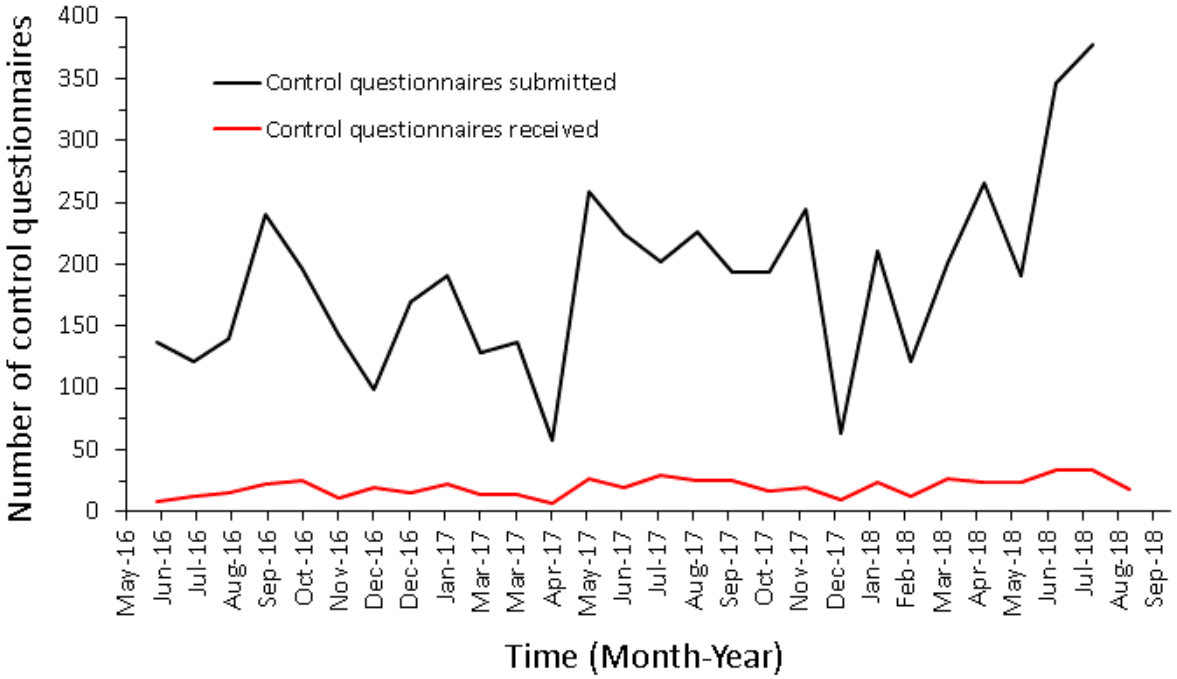
Figure 6.2 provides the temporal pattern of questionnaires sent to and returned for cases and controls. Both patterns appear similar. There is a peak in submission for both cases and controls at the end of the study whilst there is a falloff in questionnaires returned.

**Figure 6.2 Temporal pattern of questionnaires submitted and returned for (a) cases and (b) controls.**

(a)




(b)




---

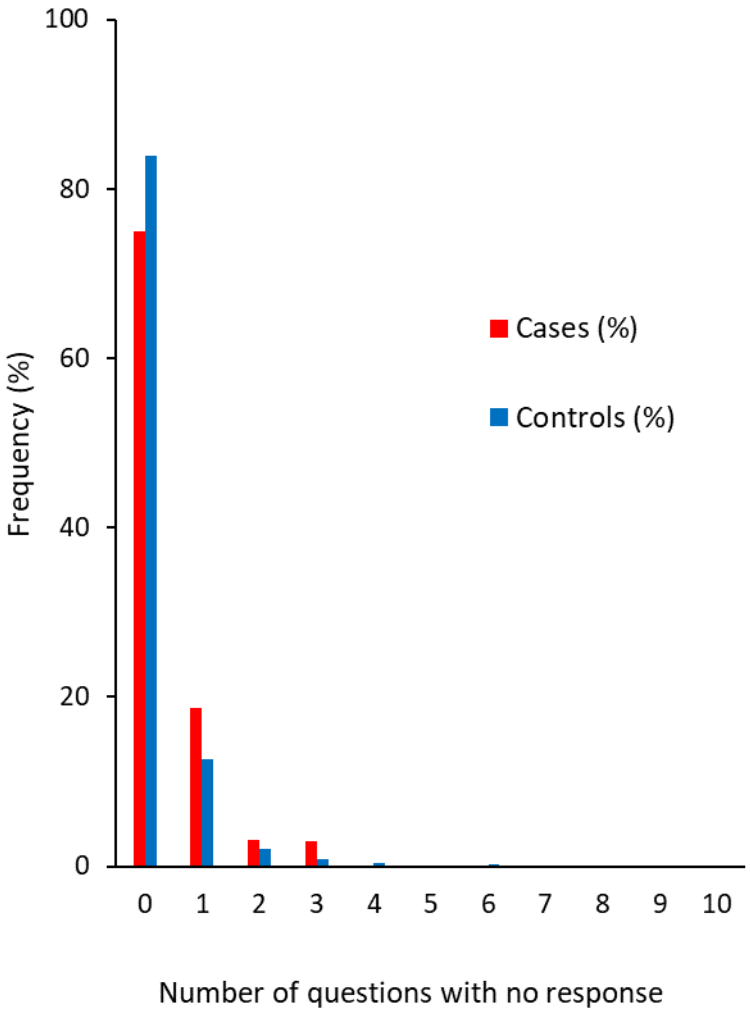
The completed questionnaires for both cases and controls were checked for quality assurance purposes. This was carried out by checking whether mandatory questions (e.g. either a closed question response or by provision of

specific information such as age) were completed. Fourteen and 12 questions from the case and control questionnaires satisfied this criteria (Table 6.4). It was found that 75% and 84% of cases and controls completed all of these questions and 94% and 97% respectively only failing to complete one (Fig. 6.3). The largest number of incomplete questions was 6 for one of the controls and most of the missing responses were in the food area of the questionnaire. It was decided that all questionnaires would be kept in the case-control study and that those with missing questions would be given the 999 code to SPSS. In the case-case study all the completed questionnaires were included from cases except those where SIMD was not available (See Fig. 6.1(b)).

**Table 6.4. The number (and percentage) of mandatory questions that were not answered by case and control participants.**

| Mandatory Questions (14 for cases and 12 for controls) | Cases<br>N (%) | Controls<br>N(%) |
|---|---|---|
| Date | 0 (0.0) | 0 (0.0) |
| Age | 1 (0.2) | 22 (4.0) |
| Sex | 0 (0.0) | 0 (0.0) |
| Postcode/SIMD | 8 (1.3) | 6 (1.1) |
| ON benefits/allowances | 10 (1.7) | 16 (2.9) |
| When first felt unwell | 20 (3.3) | NA |
| How many days after feeling unwell did you make an appointment to see the doctor? | 19 (3.2) | NA |
| Travel in Scotland | 9 (1.5) | 16 (2.9) |
| Animal contact | 3 (0.5) | 8 (1.4) |
| Water activity | 6 (1.0) | 6 (1.1) |
| Vegetarian | 0 (0.0) | 0 (0.0) |
| Eat chicken prepared at home | 34 (5.7) | 8 (1.4) |
| Eat poultry prepare at home | 35 (5.9) | 15 (2.7) |
| Eat beef, pork, lamb, deer or rabbit | 39 (6.5) | 24 (4.3) |

**Figure 6.3 Distribution of mandatory questions where there was no response from cases and controls.**




## 6.6 Determining whether there is a bias in the case and control populations responding to questionnaires (case – control analysis)

It is important to establish whether there is a bias in those returning questionnaires compared to the population to which the questionnaires were sent to. For example it may be that elderly people are more likely to return a questionnaire than young people. It is possible to correct any such biases found by weighting case and control respondents appropriately (Hosmer, Lemeshow et al. 2013).

For cases the reference population is campylobacteriosis cases reported to national surveillance >5 years old and originating from SIMD1 and SIMD5 data zones. The proportion of these cases associated with each of the following

population attributable risk factors was determined: deprivation, gender, age, season and demographic area. This was then compared with the proportion of these risk factors in the case questionnaires that were returned.

The same was performed for controls, but for them the reference population was the human population >5 years old and living in SIMD1 and SIMD5 data zones in the study area.

## 6.6.1 Cases

### 6.6.1.1 Data and Methods

Human campylobacteriosis SIMD1 and SIMD5 cases (n=2,890) reported to national surveillance for the period June 2016 to March 2018 were available. This did not encompass the whole case control study period as it went on to the end of August 2018 (i.e. only 82% (22 months /27 months)). Hence, only those SIMD1 & SIMD5 case questionnaires (n=422) returned during the same time period were also stratified by the population attributable risk factors.

Data on the following factors were used in the analysis of biases: deprivation (SIMD1 & SIMD5), gender (Male & Female), age (5-14, 15-24, 25-64 and 65+ years old), season ("summer" – May, Jun, Jul and Aug; "rest of year" – Jan, Feb, Mar, Apr, Sep, Oct, Nov and Dec) and demographic area (Rural – population density <200 people/km$^2$, peri-Urban – population density ≥200 – ‹2500 people/km$^2$ and Urban – population density ≥2500).

The proportion of case questionnaires returned for each population attributable risk factor was determined and bootstrapped 95% confidence intervals calculated. The bootstrapped CIs were calculated by randomised replacement of the case questionnaire data (n=10,000 iterations).

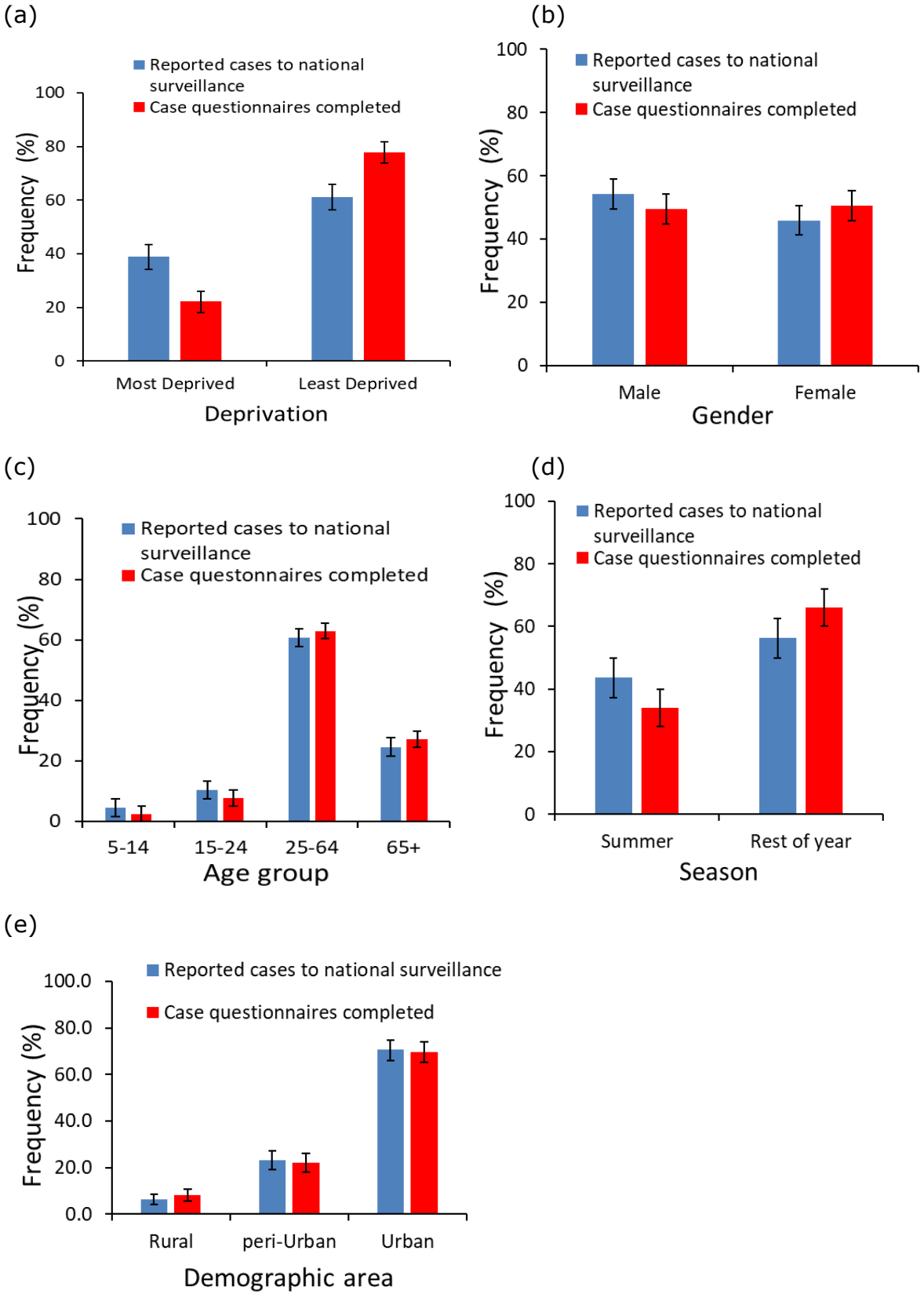
The procedure was repeated for campylobacteriosis cases reported to national surveillance. To correct for the sample size the data from national surveillance were resampled with replacement using samples of identical size as those used for the case questionnaire data. The results were plotted for each population attributable risk factor and significant differences (P-values) were estimated using randomisation tests which compares the frequency of the factor in the reported cases to national surveillance with that of the questionnaire responses (Manly 2007).

### 6.6.1.2 Results case bias

There were significant (P<0.05) biases in the return of case questionnaires by deprivation and season by the randomisation text (Figure 6.4) (Note: since bootstrapped CIs are used, on occasion they can overlap but the randomisation test can still find significance – this has happened here for season). There were fewer SIMD1 (most deprived) case questionnaires returned than expected (22% SIMD1 returned compared with 34% SIMD1 reported to national surveillance). The opposite pattern was observed for SIMD5 (less deprived) case questionnaires (Figure 6.4 (a)). With regard to season there were fewer questionnaires returned during the summer than expected (34% returned cases

during the summer compared with 44% reported cases during the same time period) (Figure 6.4 (d)). Whereas the opposite pattern was observed for the rest of the year. There were no biases for the other population attributable risk factors.

**Figure 6.4 Bias in case questionnaires returns compared with reporting of campylobacteriosis to national surveillance by population attributable risk factor.**




(a) deprivation; (b) gender; (c) age; (d) season and (e) demographic area.

## 6.6.2 Controls

### 6.6.2.1 Data and Methods

SIMD1 & SIMD5 control questionnaires (n=552) returned from the case-control study (June 2016 to August 2018) were stratified by the population attributable risk factors mentioned above in section 6.6.1.

The controls were randomly selected from the SIMD1 and SIMD5 Scottish population in the study area. Therefore the control data had to be compared with the SIMD1 & 5 population (n=1,887,283) stratified by the population attributable risk factors mentioned in section 6.6.1. The population data were obtained from SIMD (http://www.gov.scot/Topics/Statistics/SIMD/DataAnalysis/Background-Data-2012) and from the National Records of Scotland (https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates).

The proportion of control questionnaires returned for each population attributable risk factor was determined and bootstrapped 95% confidence intervals were calculated in a similar way as used for cases in section 6.6.1. These were compared with the corresponding proportions of population in the study area for each population attributable risk factor. Correction for sample sizes were applied as described in 6.1.1.1. The results from both controls and population were plotted for each population attributable risk factor and significant differences (P-values) were estimated using randomisation tests (Manly 2007).

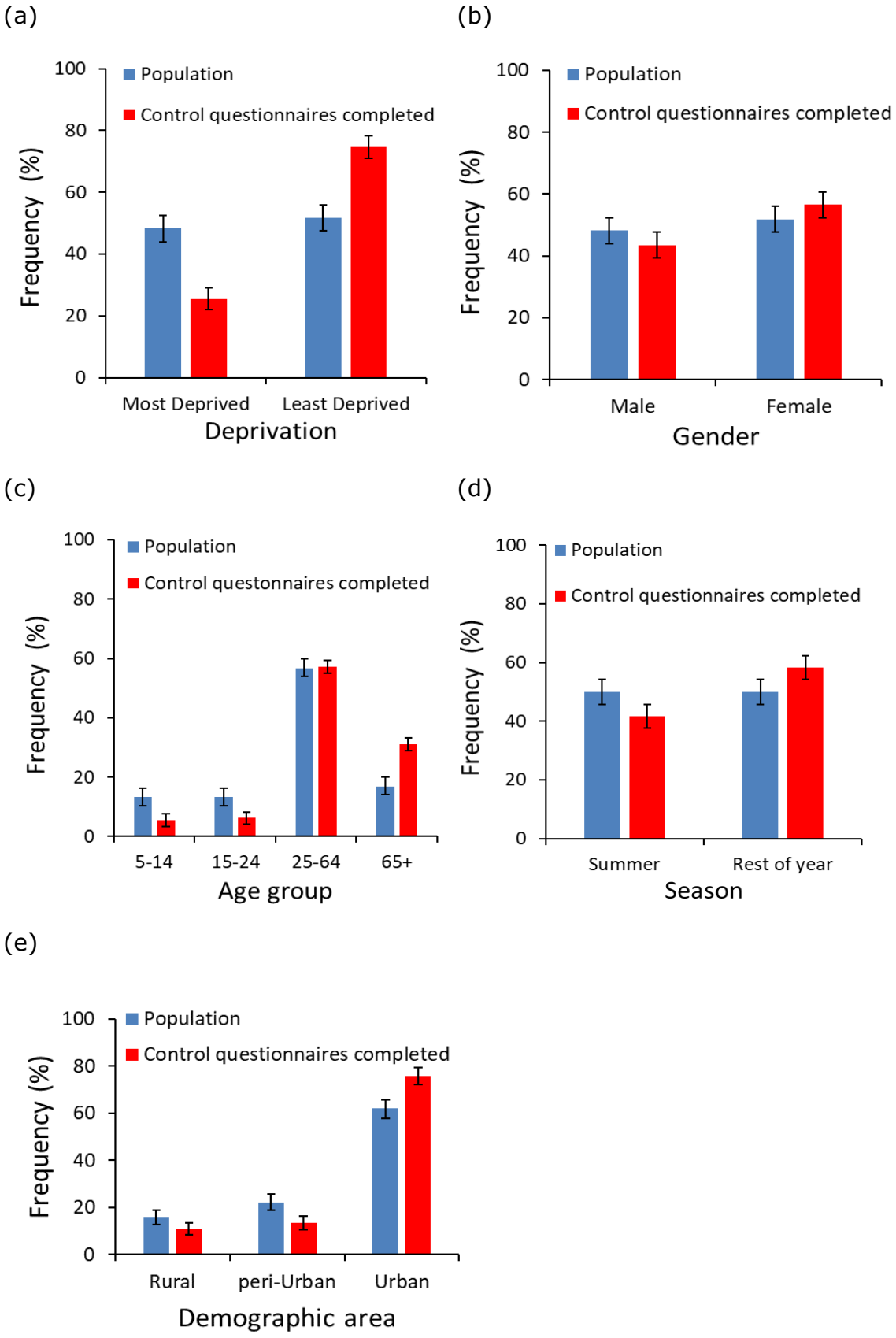
### 6.6.2.2 Results control bias

There were significant ($P<0.05$) biases in the return of control questionnaires by deprivation, age, season and demographic area (Figure 6.5). There were fewer SIMD1 (most deprived) control questionnaires returned (26%) than expected (48%) (Figure 6.5 (a)). The opposite pattern was found for SIMD5 (less deprived) control questionnaires. There were more control questionnaires than average returned from the 65+ years old age group (17% expected questionnaires vs 31% returned) (Figure 6.5 (c)). This resulted in relatively lower return rates for 5-14 years old (5.5% returned vs. 13.2% expected) and 15-24 years old (6.2% returned vs. 13.2% expected). The participants returned proportionally fewer (42%) control questionnaires during the summer (Figure 6.5 (d)) than during the rest of year. Finally, residents of urban areas returned more control questionnaires (76%) than expected (62%) (Figure 6.5 (e)) with correspondingly relatively lower rates from rural (10.8% returned vs. 15.9% expected) and peri-urban (13.4% returned vs. 22.2% expected) areas, respectively. No bias by gender was detected.

There are a number of potential explanations why these biases occurred. For example, for both cases and controls the lower return rates from the deprived population may be associated with the complexity of the questionnaire. It is known that educational attainment is lower in deprived areas (Perry, Dempster et al. 2017) and it may have been more problematic to complete the

questionnaire for those with a lower literacy. Regarding the lower return rates in the summer, it is plausible that because of the holiday period individuals are less likely to respond and in the summer because of the better weather, there is the potential that respondents are involved in other activities which leaves less time to complete the questionnaire.

**Figure 6.5 Bias in control questionnaire returns compared with the population of the study area by population attributable risk factor**




(a) deprivation; (b) gender; (c) age; (d) season and (e) demographic area.

## 6.7 Correction of bias in case and control populations responding to the Questionnaire (case – control analysis)

Reporting biases presented in sections 6.6.1.2 for cases and 6.6.2.2 for controls were used to calculate correction weights (Hosmer, Lemeshow et al. 2013), which were then used in the multivariate logistic regression case-control analysis (see Chapter 7).

### 6.7.1. Data and methods

Correction weights were calculated for the demographic factors (deprivation (SIMD1 and SIMD5), age (5-14, 15-24, 25-64 and 65+ years old), demographic area (Rural, peri-Urban and Urban)) and season (Summer and Rest of year). All of these descriptors exhibited significant bias for either cases or controls compared with the nationally reported case data (for cases) or population (for controls) respectively. The weights for the factors above were combined and applied to each individual case or control used in the multivariate logistic regression analysis (see Chapter 7).

For each of the factors mentioned above, the frequencies (%) of cases reported to the national surveillance and the frequencies (%) of cases from case questionnaire study (Figure 6.4) were used to calculate the weights for cases as follows

$$W_{Cases}^{i} = \frac{Frequency(\%)_{ReportedCases}^{i}}{Frequency(\%)_{CaseStudyQuestionnaires}^{i}}, \quad (6.1)$$

where $i$ is one of the factors mentioned above (e.g. SIMD1).

For example SIMD1 cases were weighted as follows

$$W_{Cases}^{SIMD1} = \frac{Frequency(\%)_{ReportedCases}^{SIMD1}}{Frequency(\%)_{CaseStudyQuestionnires}^{SIMD1}}. \quad (6.2)$$

Similarly, for controls the weights were calculated as follows

$$W_{Controls}^{i} = \frac{Frequency(\%)_{Population}^{i}}{Frequency(\%)_{ControlStudyQuestionnaires}^{i}}, \quad (6.3)$$

which for SIMD1 becomes

$$W_{Controls}^{SIMD1} = \frac{Frequency(\%)_{Population}^{SIMD1}}{Frequency(\%)_{ControlStudyQuestionnires}^{SIMD1}}. \quad (6.4)$$

The combined weights for cases ($W_{Cases}^{Combined}$) were simply obtained by multiplying the individual weights together for each case:

$$W_{Cases}^{Combined} = W_{Cases}^{i=1} \times W_{Cases}^{i=2} \times \ldots \quad \ldots\ldots \times W_{Cases}^{i=n} \quad (6.5)$$

where n =4 (for Season, SIMD, Rurality and Age).

For example, using the data from Table 6.5, a case during the Summer, being in a SIMD1 data zone, in a Rural area and 5-14 years old was weighted as

$$W_{Cases}^{Combined} = 1.29 \times 1.75 \times 0.77 \times 1.82 = 3.16. \quad (6.6)$$

Similarly for controls

$$W_{Controls}^{Combined} = W_{Controls}^{i=1} \times W_{Controls}^{i=2} \times \quad \ldots\ldots \times W_{Controls}^{i=n}. \qquad (6.7)$$

For example a control during the Summer, being in a SIMD1 data zone, in a Rural area and 5-14 years old was weighted as

$$W_{Controls}^{Combined} = 1.20 \times 1.9 \times 1.47 \times 2.41 = 8.08. \qquad (6.8)$$

## 6.7.2 Results and Discussion

The same weights were used for both the logistic regression of domestic case-control data and foreign travel data, respectively.

Table 6.5 presents the correction weights used in the multivariate case-control logistic regression analyses (i.e. domestic and foreign travel) for all four factors we corrected for - deprivation (SIMD1 and SIMD5), age (5-14, 15-24, 25-64 and 65+ years old),demographic area (Rural, peri-Urban and Urban) and season (Summer and rest of year).

**Table 6.5 Correction weights used in the multivariate case-control logistic regression.**

| Case - Control | Season | Weight | SIMD | Weight | Rurality | Weight | Age(y) | Weight |
|---|---|---|---|---|---|---|---|---|
| Case | Summer | 1.29 | 1 | 1.75 | Rural | 0.77 | 5-14 | 1.82 |
| | Rest of year | 0.85 | 5 | 0.79 | peri-Urban | 1.04 | 15-24 | 1.36 |
| | | | | | Urban | 1.01 | 25-64 | 0.97 |
| | | | | | | | 65+ | 0.91 |
| Control | Summer | 1.2 | 1 | 1.9 | Rural | 1.47 | 5-14 | 2.41 |
| | Rest of year | 0.86 | 5 | 0.69 | peri-Urban | 1.65 | 15-24 | 2.12 |
| | | | | | Urban | 0.82 | 25-64 | 0.99 |
| | | | | | | | 65+ | 0.54 |

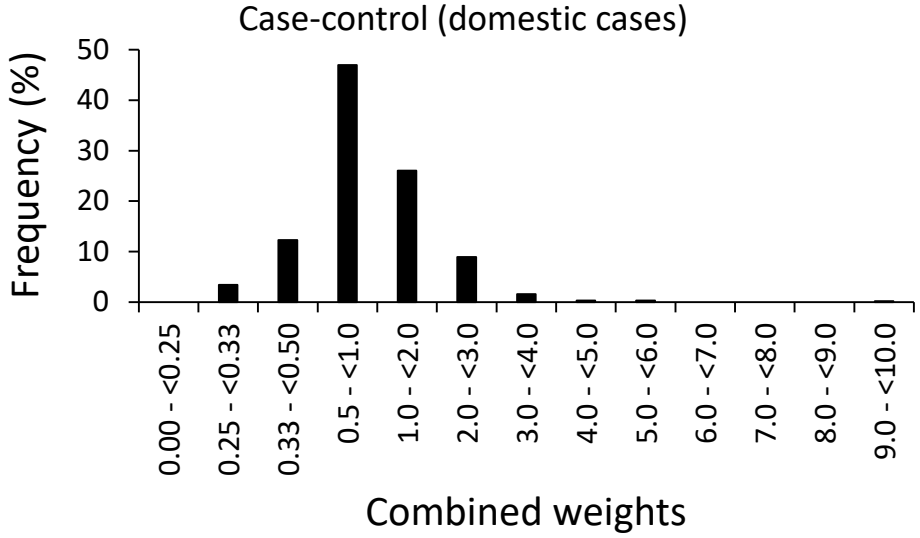
Figure 6.6(a) shows the frequency distribution of the combined weights for all domestic cases (n=452) and controls (n=500) used in the logistic regression analysis. The average weight was 1.09 (min-0.26, max-9.07).

Figure 6.6(b) shows the frequency distribution of the combined weights for all foreign travel cases (n=146) and controls (n=52) used in the logistic regression analysis. The average weight was 0.93 (min-0.32, max-3.06). Hence, using the same weights for domestic case control data and foreign travel case control datasets led to slightly different distributions. This was due to the different

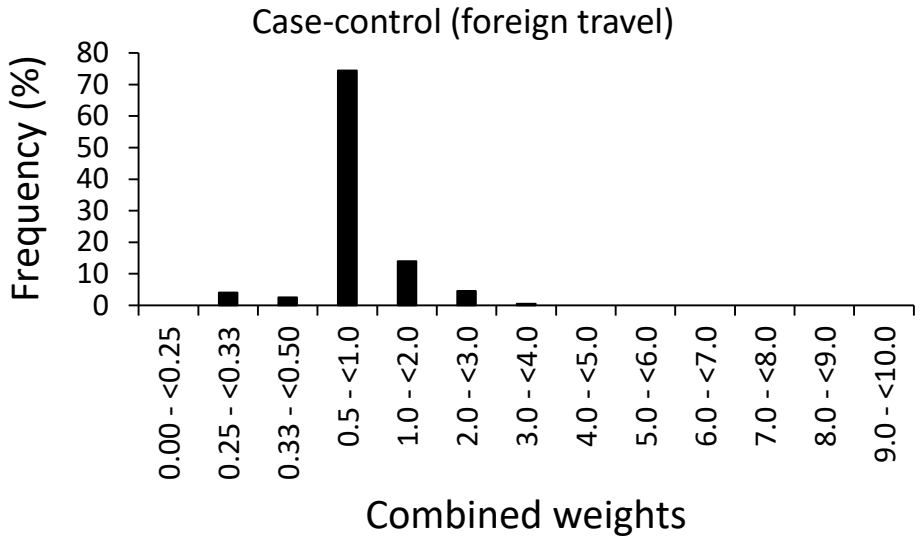
distributions of demography and seasonality between these two populations. It should be noted that cases reported to national surveillance do not capture foreign travel information. This would potentially have been a better source of data on which to weight the foreign travel cases.

**Figure 6.6 Frequency distribution of the combined weights used in the multivariate case-control logistic regression.**

(a)




(b)




(a) Domestic and (b) foreign travel associated cases and controls.

## 6.8 Determining whether there is a bias in the SIMD1 and SIMD5 cases responding to questionnaires (Case – Case analysis)

As part of objective 8, this study aims to identify differences in campylobacteriosis risk factor exposures between SIMD1 and SIMD5 populations by a case-case analysis. However, as described above for cases and controls, there may also be biases in those SIMD1 and SIMD5 populations responding to the questionnaire compared with those reported by national surveillance. This section seeks to identify such biases.

### 6.8.1 SIMD1 Cases

#### 6.8.1.1 Data and Methods

Human campylobacteriosis SIMD1 (n=1,123) reported to National Surveillance for the period June 2016 to March 2018 (i.e. 82% (22/27 months) of the case-control time period) were stratified by the population attributable risk factors. SIMD1 case questionnaires (n=92) returned during the same time period were also stratified by the population attributable risk factors.

Data on the following factors were used in the analysis of biases: deprivation gender (Male & Female), age (5-14, 15-24, 25-64 and 65+ years old), season ("summer" – May, Jun, Jul and Aug; "rest of year" – Jan, Feb, Mar, Apr, Sep, Oct, Nov and Dec) and demographic area (Rural – population density <200 people/km$^2$, peri-Urban – population density ≥200 – ‹2500 people/km$^2$ and Urban – population density ≥2500).

The proportion of SIMD1 case questionnaires returned for each population attributable risk factor was determined and bootstrapped 95% confidence intervals calculated. The bootstrapped CIs were calculated by randomised replacement of the case questionnaire data (n=10,000 iterations).

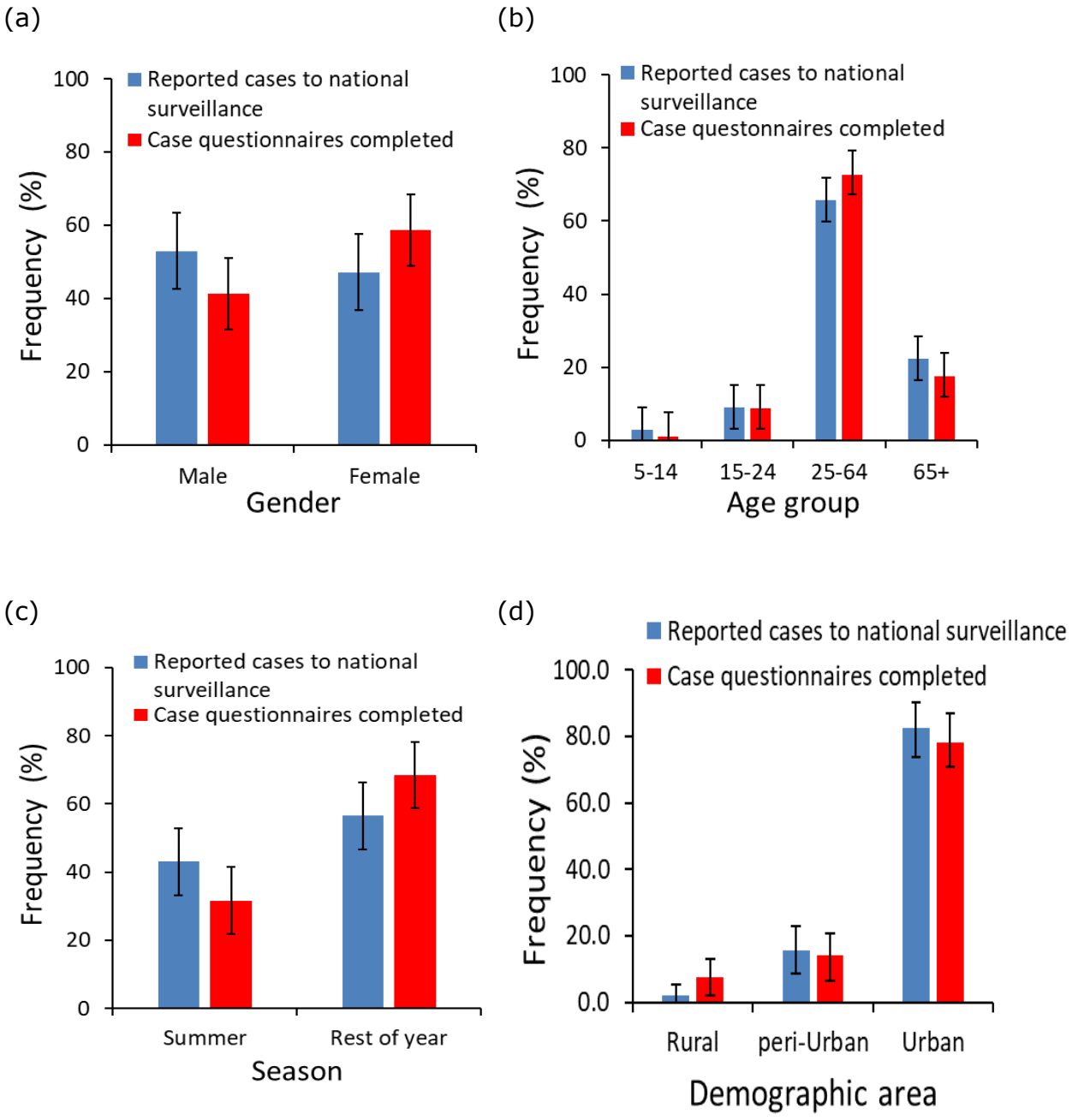
The procedure was repeated for campylobacteriosis SIMD1 cases reported to national surveillance. To correct for the sample size, the data from national surveillance were resampled with replacement using samples of identical size as those used for the case questionnaire data. The results were plotted for each population attributable risk factor and significant differences (P-values) were estimated using randomisation tests (Manly 2007).

#### 6.8.1.2 Results SIMD1 case bias

There were significant (P<0.05) biases in the return of SIMD1 case questionnaires by gender, season and demographic area (Figure 6.7). There were fewer male case questionnaires returned than expected (41% male returned compared with 53% male reported to national surveillance). The opposite pattern was observed for female case questionnaires (Figure 6.7 (a)). With regard to season there were fewer questionnaires returned during the summer than expected (32% returned cases during the summer compared with 43% reported cases during the same time period) (Figure 6.7 (c)). Whereas the opposite pattern was observed for the rest of the year. With regard to

demographic area there were more case questionnaires returned from rural areas than expected (8% rural questionnaires returned compared with 2% rural reported to national surveillance) (Figure 6.7 (d)). There were no differences by peri-Urban and Urban areas. Also, there were no biases by age groups.

**Figure 6.7 Bias in SIMD1 case questionnaires returns compared with reporting of campylobacteriosis to national surveillance by population attributable risk factor.**




(a) gender; (b) age; (c) season and (d) demographic area

## 6.8.2 SIMD5 Cases

### 6.8.2.1 Data and Methods

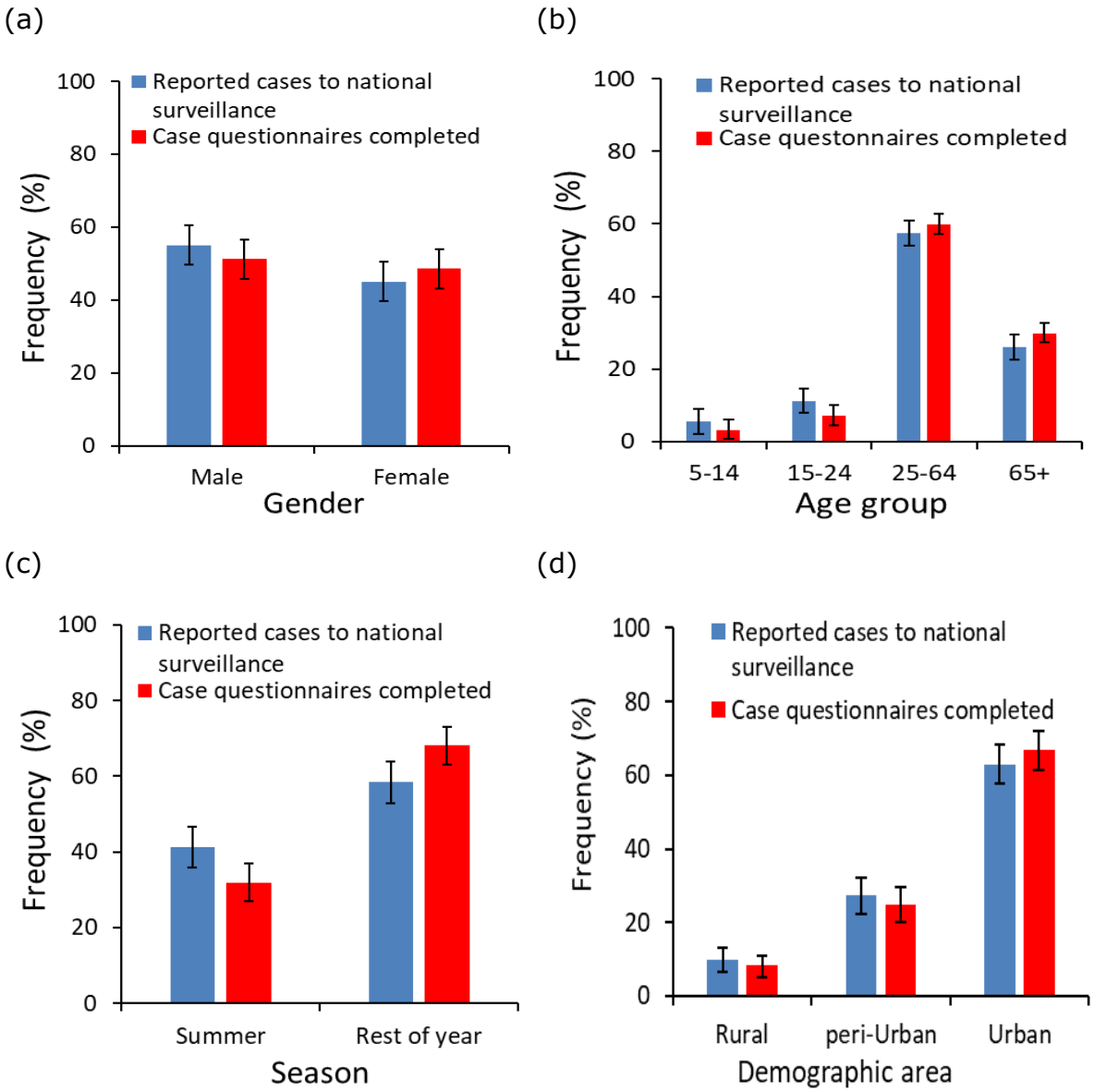
Human campylobacteriosis SIMD5 (n=1,767) reported to national surveillance for the period June 2016 to March 2018 (i.e. 82% (22/27 months) of the case-control time period) were stratified by the population attributable risk factors. SIMD5 case questionnaires (n=323) returned during the same time period were also stratified by the population attributable risk factors.

Biases using SIMD5 cases were calculated for the same factors as for SIMD1 cases and using the same approach (see section 6.7.1.1 above).

### 6.8.2.2 Results SIMD5 case bias

There were significant ($P<0.05$) biases in the return of SIMD5 case questionnaires by age and season (Figure 6.8). There were fewer case questionnaires returned by 15-24 years old than expected (3.3% of 15-24 returned compared with 5.4% of 15-24 years old reported to national surveillance). With regard to season there were fewer questionnaires returned during the summer than expected (32% returned cases during the summer compared with 42% reported cases during the same time period) (Figure 6.8 (c)). Whereas the opposite pattern was observed for the rest of the year. There were no biases for the other population attributable risk factors.

**Figure 6.8 Bias in SIMD5 case questionnaires returns compared with reporting of campylobacteriosis to national surveillance by population attributable risk factor.**




(a) gender; (b) age; (c) season and (d) demographic area.

## 6.9 Correction of bias in SIMD1 and SIMD5 populations responding to the questionnaire (case – case analysis)

Reporting biases presented in sections 6.8.1.2 for SIMD1 cases and 6.8.2.2 for SIMD5 cases were used to calculate correction weights, which were used in the multivariate logistic regression case-case analysis (see Chapter 8).

### 6.9.1 Data and methods

Correction weights were calculated for the demographic descriptors (gender (Male and Female), age (5-14, 15-24, 25-64 and 65+ years old), demographic area (Rural, peri-Urban and Urban)) and season (Summer and Rest of year). All

of these descriptors had a significant bias either for SIMD1 cases or SIMD5 cases when compared with the reported case data (see section 6.8.1.2 and 6.8.2.2). The weights for the factors above were combined and applied to each individual SIMD1 or SIMD5 case used in the multivariate logistic regression analysis of domestic cases.

For each of the descriptors mentioned above the frequencies (%) of SIMD1 cases reported to the national surveillance and the frequencies (%) of SIMD1 cases from case questionnaire study (Figure 6.7) were used to calculate the weights for the SIMD1 cases as follows:

$$W^{i}_{SIMD1Cases} = \frac{Frequency(\%)^{i}_{SIMD1ReportedCases}}{Frequency(\%)^{i}_{SIMD1CaseStudyQuestionnaires}}, \quad (6.9)$$

where $i$ is one of the factors mentioned above (e.g. Male).

For example SIMD1 male cases were weighted as follows

$$W^{Male}_{SIMD1Cases} = \frac{Frequency(\%)^{Male}_{SIMD1ReportedCases}}{Frequency(\%)^{Male}_{SIMD1CaseStudyQuestionnires}}. \quad (6.10)$$

Similarly, for SIMD5 cases the weights were calculated as follows

$$W^{i}_{SIMD5Cases} = \frac{Frequency(\%)^{i}_{SIMD5ReportedCases}}{Frequency(\%)^{i}_{SIMD5CaseStudyQuestionnaires}}, \quad (6.11)$$

which for SIMD5 male becomes

$$W^{Male}_{SIMD5Cases} = \frac{Frequency(\%)^{Male}_{SIMD5ReportedCases}}{Frequency(\%)^{Male}_{SIMD5CaseStudyQuestionnires}}. \quad (6.12)$$

The combined weights for SIMD1 cases ($W^{Combined}_{SIMD1Cases}$) were simply obtained by multiplying the individual weights together for each case:

$$W^{Combined}_{SIMD1Cases} = W^{i=1}_{SIMD1Cases} \times W^{i=2}_{SIMD1Cases} \times \ldots \quad \ldots\ldots \times W^{i=n}_{SIMD1Cases} \quad (6.13)$$

where n =4 (for Gender, Rurality, Age and Season).

For example, using data from Table 6.6, an SIMD1 case during the Summer, being Male, in a Rural area and 5-14 years old was weighted as

$$W^{Combined}_{SIMD1Cases} = 1.37 \times 1.28 \times 0.28 \times 2.71 = 1.33. \quad (6.14)$$

The combined weights for SIMD5 cases ($W^{Combined}_{SIMD5Cases}$) were simply obtained by multiplying the individual weights together for each case:

$$W^{Combined}_{SIMD5Cases} = W^{i=1}_{SIMD5Cases} \times W^{i=2}_{SIMD5Cases} \times \ldots \quad \ldots\ldots \times W^{i=n}_{SIMD5Cases} \quad (6.15)$$

where n =4 (for Gender, Rurality, Age and Season).

For example a SIMD5 case during the Summer, being Male, in a Rural area and 5-14 years old was weighted as

$$W^{Combined}_{SIMD5Cases} = 1.30 \times 1.07 \times 1.17 \times 1.68 = 2.72. \quad (6.16)$$

## 6.9.2 Results and Discussion

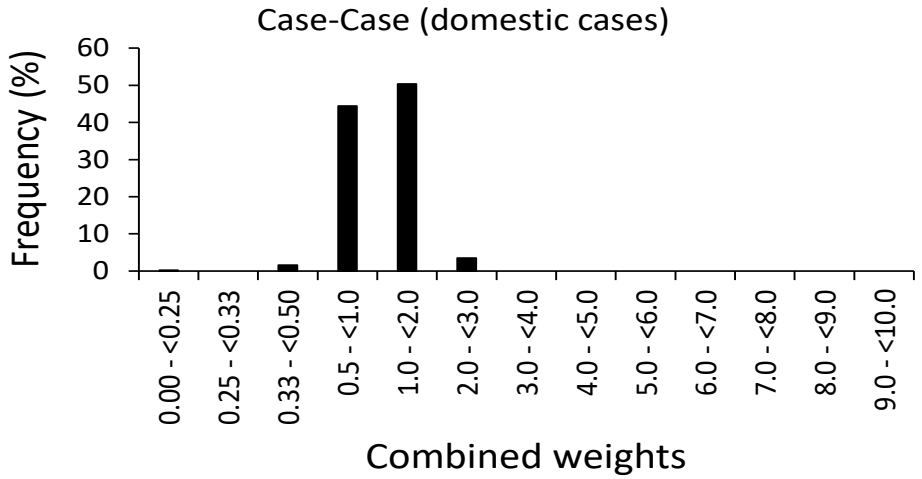
Table 6.6 presents the correction weights used in the multivariate case-case logistic regression analysis of domestic cases, for the following four correction factors - Season (Summer and Rest of year), gender (Male and Female), age (5-14, 15-24, 25-64 and 65+ years old) and demographic area (Rural, peri-Urban and Urban).

**Table 6.6 Correction weights used in the multivariate case-case logistic regression.**

| SIMD | Season | Weight | Gender | Weight | Rurality | Weight | Age(Y) | Weight |
|---|---|---|---|---|---|---|---|---|
| SIMD1 | Summer | 1.37 | Male | 1.28 | Rural | 0.28 | 5-14 | 2.71 |
| | Rest of year | 0.83 | Female | 0.80 | peri-Urban | 1.10 | 15-24 | 1.05 |
| | | | | | Urban | 1.05 | 25-64 | 0.90 |
| | | | | | | | 65+ | 1.28 |
| SIMD5 | Summer | 1.30 | Male | 1.07 | Rural | 1.17 | 5-14 | 1.68 |
| | Rest of year | 0.86 | Female | 0.93 | peri-Urban | 1.10 | 15-24 | 1.56 |
| | | | | | Urban | 0.94 | 25-64 | 0.96 |
| | | | | | | | 65+ | 0.87 |

Figure 6.9 shows the frequency distribution of the combined weights for the domestic cases (SIMD1, n=113 and SIMD5, n=332) used in the logistic case-case regression analysis. The average weight was 1.07 (min-0.17, max-2.56).

**Figure 6.9 Frequency distribution of the combined weights used in the multivariate case-case logistic regression of domestic cases.**




## 6.10 Conclusions

In total 598 cases and 552 controls were recruited for the study. The response rates were approximately 23% for cases and 11% for controls. This was lower than previous studies conducted in Scotland and elsewhere. Response rates from the most deprived SIMD quintile were lower than the least deprived for both cases and controls. Future studies could consider telephone and in person interviews as methods for improving response rates.

Overall the questionnaires from both cases and controls were completed thoroughly with <5% failing to answer more than one of the mandatory questions. Biases were observed in both case and control questionnaire responses compared with the reference population. Biases were also observed in the SIMD1 and SIMD5 questionnaire responders compared with national surveillance. This has the potential to affect both the case-case and case-control findings. Weights were calculated to correct for these biases and will be used in the logistic regression analysis for the case-control (Chapter 7) and case-case (Chapter 8) studies.

# 7. Case-control Study

## 7.1 Introduction

This chapter analyses the case-control study questionnaire data using logistic regression methods (Hosmer, Lemeshow et al. 2013). The purpose of logistic regression is the same as other regression methods and that is to find a model that is parsimonious, clinically interpretable, best fitting that explains the relationship between the outcome variable (*Campylobacter* case or control) and the explanatory variable(s) (e.g. ate undercooked chicken, contact with animals etc.). As described in the previous chapter (section 6.3) the analysis is split into domestically acquired and foreign travel associated case-control studies. The rationale for doing this is twofold. First to identify the role that deprivation (or affluence) plays for these two groups of cases. Second, to be able to identify the importance of domestic food based risk factors that is a primary focus for Food Standards Scotland. Weights are also applied to the datasets to correct for any bias in those deciding to return the questionnaires.

Initially univariate analysis was performed on all the variables extracted from the case and control questionnaires to identify those factors that are associated with increased or decreased risk of campylobacteriosis. Multivariate regression models were then built based on an appropriate selection of variables from the univariate analysis. For those risk factors that are statistically significant in the final multivariate model, the population attributable fraction is determined (i.e. the proportion of disease risk in a population that can be attributed to the causal effects of a risk factor (Miettinen 1974)).

A further analysis was performed to quantify the proportion of reported case difference between SIMD5 and SIMD1 areas (See Fig. 4.2) attributable to differences in amount of foreign travel.

It is worth noting that the case-control study by its nature only considers reported cases. Any cases that go unreported, for whatever reason (e.g. access to healthcare facilities etc.), at any point in the reporting pyramid cannot be included.

## 7.2 Perform case-control analysis using logistic regression

### 7.2.1 Data

The data extracted from the case and control questionnaires were used in this study. This included general details about the individual including, for example, age, details of household income, historical health conditions, travel and exposure to animals, food and water. The data were split into domestic and foreign travel associated cases and controls.

### 7.2.2 Methods

#### 7.2.2.1 Descriptive analysis

For each factor the number of cases and number of controls exposed were determined as well as the number of cases and controls where data were incomplete.

#### 7.2.2.2 Univariate and multivariate logistic regression

Univariate regression adjustments: For the domestic case-control analysis the following confounding variables (season, age, sex, SIMD and rurality) which had been previously identified as risk factors for human campylobacteriosis (Kuhn, Nielsen et al. 2018) were assessed by univariate logistic regression (SPSS 25) to determine whether they were significant ($P<0.05$) risk factors. For those variables that were statistically significant adjustments were made in both the univariate and multivariate analysis. This was repeated for the foreign travel case-control study but rurality was omitted as it was not considered to be an important determinant for foreign travel associated campylobacteriosis (Kuhn, Nielsen et al. 2018).

Univariate and multivariate regression domestic case-control study: The univariate analysis was performed for all explanatory variables (putative risk factors) utilising logistic regression which generated ORs and 95% confidence intervals. Variables with a p-value of $< 0.25$ were selected for the multivariate analyses. A 'relaxed' p-value of 0.25 was used as a more stringent setting of p can fail in inclusion of variables known to be important (Bursac, Gauss et al. 2008). Multivariate logistic regression was performed by backwards stepwise elimination with non-significant variables removed one step at a time. Missing data were inferred by multiple imputation with 100 iterations and a pooled model was generated. Backwards step elimination was repeated until only variables were left with $P<0.157$ and $P<0.05$. These two models were kept for further analysis.

Multivariate models were performed where the data were both unweighted and weighted to correct for sample bias (see chapter 6 sections 6.6 and 6.7).

The goodness of fit of the multivariate models can be assessed in a number of ways (e.g. the omnibus test and Hosmer-Lemeshow test (Hosmer, Lemeshow et al. 2013)). However, since analysis performed here utilises multiple imputations and a pooled model is generated it is not possible to use such tests. Therefore each model was tested to determine how many of the cases and controls were correctly assigned.

The population attributable fraction, which is the proportion of disease risk in a population that can be attributed to the causal effects of a risk factor or set of risk factors (Greenland, Robins 1988) is defined by (Miettinen 1974):

$$PAF = P_C\left(\frac{RR-1}{RR}\right)$$

Where $P_C$ is the proportion of cases exposed to the risk factor and $RR$ is the relative risk. $RR$ cannot be obtained directly from the logistic regression but the adjusted odds ratio can be used instead. It should be noted that the case-control dataset comprises a population from SIMD1 and SIMD5 only and hence the calculated $PAF$ is for that part of the Scottish population only. Confidence

intervals for the PAF were obtained by propagating through the errors in the odds-ratio.

Univariate and multivariate regression foreign travel case-control study: This was carried out as for the domestic case-control analysis. However, an additional univariate logistic regression analysis was performed solely looking at the region of destination. This comprised Africa, Asia, Australasia (Australia and New Zealand), North America, South America and Europe. Europe was split into four regions (Mughini-Gras, Smid et al. 2014): Western (Germany, France, Belgium, Austria, Luxembourg, Switzerland, The Netherlands); Eastern (Czech Republic, Hungary, Poland, Slovakia, Romania, Bulgaria); Northern (Ireland, Denmark, Sweden, Norway, Finland) and Southern (Spain, Italy, Portugal, Greece, Croatia, Malta). Univariate logistic regression was performed to generate adjusted odds ratios, confidence intervals and P values for each of the regions visited.

## 7.2.3 Results – Domestic Case-Control

### 7.2.3.1 Domestic Case-control logistic regression analysis

Table 7.1 presents the results of the univariate analysis for those variables selected as possible adjustments for confounding. There was significantly higher odds ratios for cases from the summer and male and significantly lower from those aged 5-14, 25-64 and from rural areas. As such it was decided to use season, sex, age and rurality as adjustment factors in the subsequent univariate analysis.

Deprivation and Domestic Case-Control Study: Table 7.1 shows that there is proportionally more cases from least deprived (SIMD5) than from most deprived (SIMD1) areas but that this is not statistically significant (OR=1.088, P=0.572). This is an agreement with the previous finding that there is an excess of reported cases in the least deprived population (Chapter 4). This was repeated by weighting the data to correct for differential response rate in returning questionnaires since it is known that the response rate from controls from deprived areas was very low. Although the OR increased (OR =1.217, 95% CI (0.949-1.560)), indicating a stronger relationship, it was still not statistically significant (P=0.122).

Deprivation and Foreign travel associated Case-Control Study: there appeared to be proportionally more cases from most deprived compared with least deprived areas though this was not statistically significant (Table 7.1). This is counter to what would be expected, since the previous geography study had hypothesised that part of the explanation of reduced campylobacteriosis cases in deprived areas may be due to reduced foreign travel.

**Table 7.1 Univariate analysis of potential adjusting variables overall and then for domestic and foreign travel associated cases and controls separately.**

| **Domestic -Case-control Study** | | | | | | |
|---|---|---|---|---|---|---|
| Characteristic | Cases (N=452)<br>n (%) | Cases Unk.<br>n | Controls (N=500)<br>n (%) | Controls Unknown<br>n | OR (95% CI) | P-value |
| Season | | | | | | |
| Summer | 216 (47.8) | 0 | 205 (41.0) | 0 | 1.317 (1.019 – 1.702) | 0.035 |
| Rest of Year (Ref.) | 236 (52.2) | 0 | 295 (59.0) | 0 | 1 | |
| Age | | | | | | |
| 5 - 14 | 10 (2.2) | 19 | 27 (5.4) | 22 | 0.424 (0.198 – 0.909) | 0.027 |
| 15 - 24 | 27 (6.0) | 19 | 30 (6.0) | 22 | 1.031 (0.583 – 1.823) | 0.918 |
| 25 - 64 | 265 (58.6) | 19 | 271 (54.2) | 22 | 1.120 (0.839 – 1.495) | 0.443 |
| 65+ (Ref.) | 131 (29.0) | 19 | 150 (30.0) | 22 | 1 | |
| Sex | | | | | | |
| Male | 237 (52.4) | 0 | 211 (42.2) | 0 | 1.510 (1.169 – 1.950) | 0.002 |
| Female (Ref.) | 215 (47.6) | 0 | 289 (57.8) | 0 | 1 | |
| SIMD5 | 332 (73.5) | 7 | 362 (72.4) | 4 | 1.088 (0.813 – 1.455) | 0.572 |
| SIMD1 (Ref.) | 113 (25) | 7 | 134 (26.8) | 4 | 1 | |
| Rurality | | | | | | |
| Rural | 31 (6.9) | 6 | 55 (11.0) | 4 | 0.597 (0.374 – 0.952) | 0.030 |
| Peri-urban | 108 (23.9) | 6 | 116 (23.2) | 4 | 0.986 (0.727 – 1.337) | 0.926 |
| Urban (Ref.) | 307 (67.9) | 6 | 325 (65.0) | 4 | 1 | |

| **Foreign Travel Associated Case-Control Study[a]** | | | | | | |
|---|---|---|---|---|---|---|
| | Cases (N=146)<br>n (%) | Cases Unk.<br>n | Controls (N=52)<br>n (%) | Controls Unknown<br>n | OR (95% CI) | P-value |
| Season | | | | | | |
| Summer | 62 (42.5) | 0 | 25 (48.1) | 0 | 0.797 (0.422 – 1.505) | 0.484 |
| Rest of Year (Ref.) | 84 (57.5) | 0 | 27 (51.9) | 0 | 1 | |
| Age | | | | | | |
| 5 - 14 | 3 (2.1) | 3 | 2 (3.8) | 0 | 0.978 (0.146 – 6.565) | 0.982 |
| 15 - 24 | 9 (6.2) | 3 | 3 (5.8) | 0 | 1.957 (0.455 – 8.421) | 0.367 |
| 25 - 64 | 108 (74.0) | 3 | 32 (61.5) | 0 | 2.201 (1.029 – 4.710) | 0.042 |
| 65+ (Ref.) | 23 (15.8) | 3 | 15 (28.8) | 0 | 1 | |
| Sex | | | | | | |
| Male | 56 (38.4) | 0 | 29 (55.8) | 0 | 0.493 (0.260 – 0.937) | 0.031 |
| Female (Ref.) | 90 (61.6) | 0 | 23 (44.2) | 0 | 1 | |
| | | | | | | |
| SIMD5 | 123 (84.2) | 1 | 45 (86.5) | 2 | 0.621 (0.222 - 1.739) | 0.365 |
| SIMD1 (Reference) | 22 (15.1)) | 1 | 5 (9.6) | 2 | 1 | |

a Foreign travel associated cases were not analysed in terms of rurality of home address as detailed in Kuhn et al., 2019.

Table 7.2 provides the univariate logistic regression analysis. Tables 7.3 and 7.4 provide the results for the unweighted and weighted multivariate logistic regression analysis with P<0.05. The corresponding models with P<0.157 are provided in Annex 7.1 (Tables A7.1 and A7.2). It can be seen from Table 7.2 that there are 15 factors that are significantly associated with campylobacteriosis risk and 15 factors that are significantly "protective". In the unweighted multivariate analysis this reduces to 8 that are significantly associated with campylobacteriosis risk and 4 that are significantly "protective" using the P<0.05 model (Table 7.3). This reduces further with the P<0.157 model (Table A7.1) to 7 that are significantly associated with campylobacteriosis risk and 1 that is protective. Most of those factors that are significant in the

P<0.05 model but not in the P<0.157 model are contained in the P<0.157 model but their P values are between 0.05 and 0.157.

Both the weighted and unweighted multivariate P<0.05 regression models (Tables 7.3 and 7.4) had 8 factors that were significantly associated with campylobacteriosis risk and 4 that were protective. Nine of the factors were the same in both models. The three that were different in each model can be considered as pairs. The first relates to pre-existing illness where long term bowel illness is a risk factor in the unweighted model whilst usage of antiflatuents is a risk factor in the weighted model. The second is animal contact where in the unweighted model contact with animals is "protective" whist in the weighted model a subset of this group contact with cats is "protective. Finally the third deals with consumption of other foods. In the unweighted model eating pork oven roasted or grilled is "protective" whilst in the weighted model eating beef oven cooked, roasted or grilled is "protective". It may be that these foods are of lower risk of containing *Campylobacter* and so this may be being eaten more than other foods of higher risk and hence come out "protective."

Table 7.5 shows the percentage of cases and controls correctly assigned in the final multivariate models. It can be seen that the weighted and unweighted models produce very similar assignments. The P<0.157 models are slightly better than the P=0.05 models (by approx. 1%) but this is not surprising as the P<0.157 models comprise more variables.

The tables are presented below and there follows a specific discussion of the risk factors.

**Table 7.2 Univariate analysis of Domestic Risk Factors**

| General Details – Personal Characteristics | Cases, n (%) | Cases Unk.[a] | Controls, n (%) | Controls Unk.[a] | Adjusted OR (95% CI) | P-value |
|---|---|---|---|---|---|---|
| N | 452 | | 500 | | | |
| **Season** | | | | | | |
| Summer | 216 (47.8) | 0 | 205 (41.0) | 0 | 1.33 (1.01-1.73) | 0.040 |
| Rest of year (Reference) | 236 (52.2) | 0 | 295 (59.0) | 0 | | |
| **Age (years)** | | | | | | |
| 5-14 | 10 (2.2) | 19 | 27 (5.4) | 22 | 0.43 (0.20-0.93) | 0.033 |
| 15-24 | 27 (6.0) | 19 | 30 (6.0) | 22 | 1.26 (0.69-2.28) | 0.452 |
| 25-64 | 265 (58.6) | 19 | 271 (54.2) | 22 | 1.20 (0.89-1.62) | 0.223 |
| 65+ (Reference) | 131 (29.0) | 19 | 150 (30.0) | 22 | | |
| **Sex** | | | | | | |
| Male | 237 (52.4) | 0 | 211 (42.2) | 0 | 1.60 (1.23-2.1) | 0.001 |
| Female (Reference) | 215 (47.6) | 0 | 289 (57.8) | 0 | | |
| **Ethnicity** | | | | | | |
| White | 447 (98.9) | 0 | 489 (97.8) | 0 | 2.48 (0.84-7.31) | 0.101 |
| Other (Reference) | 5 (1.1) | 0 | 11 (2.2) | 0 | | |

| (continued) | Cases, n (%) | Cases Unk. | Controls, n (%) | Controls Unk. | Adjusted OR (95% CI) | P-value |
|---|---|---|---|---|---|---|
| **SIMD** | | | | | | |
| 5 | 332 (73.5) | 7 | 362 (72.4) | 4 | 1.10 (0.81-1.5) | 0.531 |
| 1 (Reference) | 113 (25) | 7 | 134 (26.8) | 4 | | |
| **Rurality, benefits and income** | | | | | | |
| Rural | 31 (6.9) | 6 | 55 (11.0) | 4 | 0.60 (0.37-0.97) | 0.038 |
| peri-Urban | 108 (23.9) | 6 | 116 (23.2) | 4 | 1.00 (0.73-1.38) | 0.978 |
| Urban (Reference) | 307 (67.9) | 6 | 325 (65.0) | 4 | | |
| Benefits | 66 (14.6) | 10 | 60 (12) | 16 | 1.25 (0.84-1.85) | 0.270 |
| Household income >£47k/year | 159 (35.2) | 25 | 190 (38) | 22 | 0.93 (0.70-1.25) | 0.631 |
| **Occupation** | | | | | | |
| Retired | 159 (35.2) | 0 | 163 (32.6) | 0 | 1.24 (0.85-1.82) | 0.259 |
| School aged | 16 (3.5) | 0 | 31 (6.2) | 0 | 0.61 (0.77-0.29) | 0.606 |
| Student | 18 (4.0) | 0 | 27 (5.4) | 0 | 0.56 (0.26-1.21) | 0.140 |
| Unemployed | 22 (4.9) | 0 | 31 (6.2) | 0 | 0.72 (0.39-1.31) | 0.282 |
| Professional job | 163 (36.1) | 0 | 172 (34.4) | 0 | 1.05 (0.77-1.45) | 0.751 |
| Car <5 years old | 229 (50.7) | 0 | 209 (41.8) | 0 | 1.41 (1.07-1.85) | 0.013 |
| House >= 3 bedrooms | 293 (64.8) | 0 | 318 (63.6) | 0 | 1.09 (0.82-1.44) | 0.551 |

| (continued) | Cases, n (%) | Cases Unk. | Controls, n (%) | Controls Unk. | Adjusted OR (95% CI) | P-value |
|---|---|---|---|---|---|---|
| Sit/Sat on a committee or council | 60 (13.3) | 0 | 79 (15.8) | 0 | 0.81 (0.55-1.20) | 0.302 |
| Profession - None of the above | 42 (9.3) | 0 | 45 (9.0) | 0 | 1.05 (0.67-1.66) | 0.822 |
| **Historical health conditions and treatment** | | | | | | |
| Long term bowel condition | 78 (17.3) | 0 | 47 (9.4) | 0 | 1.89 (1.26-2.83) | 0.002 |
| Other medical condition | 180 (39.8) | 0 | 181 (36.2) | 0 | 1.10 (0.83-1.46) | 0.489 |
| PPIs | 106 (23.5) | 0 | 52 (10.4) | 0 | 2.42 (1.65-3.54) | 0.001 |
| H2-blockers | 12 (2.7) | 0 | 9 (1.8) | 0 | 1.40 (0.58-3.40) | 0.456 |
| Antacids | 54 (11.9) | 0 | 65 (13.0) | 0 | 0.84 (0.56-1.25) | 0.388 |
| Antiflatuents | 12 (2.7) | 0 | 7 (1.4) | 0 | 2.05 (0.79-5.33) | 0.140 |
| PPIs, H2-blockers, Antacids, Antiflatuents | 157 (34.7) | 0 | 112 (22.4) | 0 | 1.65 (1.22-2.23) | 0.001 |
| Antibiotics | 29 (6.4) | 0 | 27 (5.4) | 0 | 1.19 (0.69-2.06) | 0.537 |
| Medicine (other) | 202 (44.6) | 0 | 213 (42.6) | 0 | 1.13 (0.86-1.48) | 0.382 |
| **Travel** | | | | | | |
| Travel within Scotland | 80 (17.7) | 7 | 105 (21.0) | 15 | 0.84 (0.60-1.18) | 0.311 |
| Travel outside Scotland (not abroad) | 40 (8.8) | 42 | 45 (9.0) | 34 | 1.01 (0.65-1.58) | 0.960 |
| Foreign travel | 0 (0) | 46 | 0 (0) | 36 | nd | nd |

| (continued) | Cases, n (%) | Cases Unk. | Controls, n (%) | Controls Unk. | Adjusted OR (95% CI) | P-value |
|---|---|---|---|---|---|---|
| **Contact with animals** | | | | | | |
| Contact with animals -overall | 228 (50.4) | 3 | 291 (58.2) | 8 | 0.75 (0.57-0.99) | 0.044 |
| Dogs | 172 (38.1) | 0 | 223 (44.6) | 0 | 0.81 (0.61-1.06) | 0.126 |
| Cats | 73 (16.2) | 0 | 103 (20.6) | 0 | 0.79 (0.56-1.12) | 0.185 |
| Birds/Poultry | 16 (3.5) | 0 | 14 (2.8) | 0 | 1.36 (0.64-2.90) | 0.428 |
| Farm animals(cattle, sheep, goats, horses, donkeys, pigs) | 10 (2.2) | 0 | 16 (3.2) | 0 | 0.76 (0.33-1.74) | 0.521 |
| Other animals | 18 (4.0) | 0 | 36 (7.2) | 0 | 0.47 (0.25-0.88) | 0.019 |
| Contact with ill animal | 11 (2.4) | 0 | 9 (1.8) | 0 | 1.46 (0.58-3.68) | 0.427 |
| Touch animal faeces | 38 (8.4) | 42 | 59 (11.8) | 25 | 0.76 (0.48-1.21) | 0.249 |
| **Exposure to water** | | | | | | |
| Water activity | 41 (9.1) | 5 | 73 (14.6) | 6 | 0.55 (0.35-0.84) | 0.006 |
| Indoor swimming pool / toddler pool | 25 (5.5) | 5 | 59 (11.8) | 6 | 0.44 (0.26-0.73) | 0.002 |
| Outdoor swimming pool / paddling pool / theme park water ride / splash park | 2 (0.4) | 5 | 1 (0.2) | 6 | 1.22 (0.07-20.14) | 0.888 |
| Loch/lake/pond/stream/river/burn (e.g. swimming, canoeing, diving, fishing) | 6 (1.3) | 5 | 9 (1.8) | 6 | 0.75 (0.26-2.17) | 0.598 |
| Sea (e.g. diving, sailing, surfing, jet ski, fishing) | 3 (0.7) | 5 | 8 (1.6) | 6 | 0.32 (0.08-1.27) | 0.105 |

| (continued) | Cases, n (%) | Cases Unk. | Controls, n (%) | Controls Unk. | Adjusted OR (95% CI) | P-value |
|---|---|---|---|---|---|---|
| Other water activity | 9 (2.0) | 5 | 1 (0.2) | 6 | 6.59 (0.81-53.33) | 0.077 |
| Water source (public mains) | 426 (94.2) | 6 | 472 (94.4) | 9 | 0.72 (0.37-1.42) | 0.348 |
| Water source (private-spring) | 4 (0.9) | 6 | 9 (1.8) | 9 | 0.60 (0.18-2.00) | 0.408 |
| Water source (private-well) | 5 (1.1) | 6 | 6 (1.2) | 9 | 0.81 (0.23-2.88) | 0.746 |
| Water source (River/stream/lake/loch/pond/melted snow (not boiled) | 3 (0.7) | 6 | 1 (0.2) | 9 | 4.03 (0.39-41.85) | 0.243 |
| **Exposure to food** | | | | | | |
| **Foods eaten - chicken** | | | | | | |
| Eat chicken prepared at home | 265 (58.6) | 25 | 350 (70.0) | 7 | 0.70 (0.53-0.94) | 0.016 |
| Eat chicken outside the home | 148 (32.7) | 0 | 160 (32.0) | 0 | 1.05 (0.79-1.40) | 0.743 |
| Chicken outside the home - Restaurant | 61 (13.5) | 0 | 82 (16.4) | 0 | 0.83 (0.57-1.21) | 0.342 |
| Chicken outside the home - Take away or Fast food | 67 (14.8) | 0 | 66 (13.2) | 0 | 1.05 (0.71-1.55) | 0.807 |
| Chicken outside the home - Elsewhere | 85 (18.8) | 0 | 65 (13.0) | 0 | 1.6 (1.10-2.32) | 0.013 |
| Eat chicken liver pâté prepared from raw at home | 31 (6.9) | 35 | 10 (2.0) | 14 | 4.16 (1.93-8.99) | 0.001 |
| Eat chicken liver pâté prepared outside the home | 6 (1.3) | 0 | 5 (1.0) | 0 | 1.57 (0.46-5.29) | 0.469 |

| (continued) | Cases, n (%) | Cases Unk. | Controls, n (%) | Controls Unk. | Adjusted OR (95% CI) | P-value |
|---|---|---|---|---|---|---|
| Frozen chicken purchased which was then prepared at home | 88 (19.5) | 47 | 103 (20.6) | 36 | 0.93 (0.66-1.30) | 0.660 |
| Fresh raw chicken purchased which was then prepared at home | 244 (54.0) | 31 | 308 (61.6) | 15 | 1.18 (0.89-1.56) | 0.262 |
| Raw chicken washed before preparation | 61 (13.5) | 54 | 65 (13.0) | 27 | 1.03 (0.70-1.53) | 0.865 |
| Raw chicken cut up in the kitchen | 162 (35.8) | 54 | 228 (45.6) | 27 | 0.71 (0.53-0.96) | 0.023 |
| Raw chicken handled in the kitchen | 146 (32.3) | 55 | 220 (44.0) | 31 | 0.64 (0.48-0.86) | 0.003 |
| Raw chicken at home - oven-cooked, roasted or grilled | 137 (30.3) | 48 | 198 (39.6) | 21 | 0.83 (0.62-1.10) | 0.191 |
| Chicken outside the home - oven-cooked, roasted or grilled | 93 (20.6) | 0 | 97 (19.4) | 0 | 1.07 (0.76-1.49) | 0.710 |
| Raw chicken at home - BBQ | 19 (4.2) | 48 | 6 (1.2) | 21 | 3.16 (1.21-8.28) | 0.019 |
| Chicken outside the home - BBQ | 17 (3.8) | 0 | 15 (3.0) | 0 | 1.01 (0.48-2.11) | 0.980 |
| Raw chicken at home - stir fried | 70 (15.5) | 48 | 122 (24.4) | 21 | 0.61 (0.43-0.87) | 0.006 |
| Chicken outside the home - stir fried | 20 (4.4) | 0 | 27 (5.4) | 0 | 0.77 (0.41-1.44) | 0.419 |
| Raw chicken at home - microwaved | 2 (0.4) | 48 | 3 (0.6) | 21 | 0.84 (0.14-5.12) | 0.849 |
| Chicken outside the home - microwaved | 7 (1.5) | 0 | 5 (1.0) | 0 | 1.87 (0.54-6.49) | 0.326 |
| Raw chicken at home - stewed, slow cooked or steamed | 31 (6.9) | 48 | 48 (9.6) | 21 | 0.70 (0.42-1.15) | 0.156 |

| (continued) | **Cases, n (%)** | **Cases Unk.** | **Controls, n (%)** | **Controls Unk.** | **Adjusted OR (95% CI)** | **P-value** |
|---|---|---|---|---|---|---|
| Chicken outside the home - stewed, slow cooked or steamed | 14 (3.1) | 0 | 18 (3.6) | 0 | 0.78 (0.36-1.69) | 0.781 |
| Raw chicken at home - deep fried | 3 (0.7) | 48 | 2 (0.4) | 21 | 2.10 (0.33-13.23) | 0.428 |
| Chicken outside the home - deep fried | 22 (4.9) | 0 | 26 (5.2) | 0 | 1.07 (0.58-2.00) | 0.821 |
| Chicken lightly cooked (i.e. pinkish in the middle) | 11 (2.4) | 96 | 4 (0.8) | 21 | 4.43 (1.31-14.98) | 0.017 |
| **Foods eaten – poultry other than chicken** | | | | | | |
| Eat poultry other than chicken prepared at home | 28 (6.1) | 29 | 47 (9.4) | 14 | 0.75 (0.45-1.24) | 0.257 |
| Eat poultry other than chicken prepared outside the home | 35 (7.7) | 0 | 18 (3.6) | 0 | 2.5 (1.35-4.63) | 0.004 |
| Poultry (not chicken) outside the home - Restaurant | 25 (5.5) | 0 | 5 (1.0) | 0 | 5.58 (2.08-14.94) | 0.001 |
| Poultry (not chicken) outside the home - Take away or Fast food | 5 (1.1) | 0 | 10 (2.0) | 0 | 0.68 (0.22-2.14) | 0.515 |
| Poultry (not chicken) outside the home - Elsewhere | 20 (4.4) | 0 | 10 (2.0) | 0 | 2.70 (1.20-6.05) | 0.016 |
| Eat poultry liver pâté (not chicken) which was prepared from raw at home | 4 (0.9) | 33 | 2 (0.4) | 18 | 1.94 (0.31-12.18) | 0.481 |

| (continued) | Cases, n (%) | Cases Unk. | Controls, n (%) | Controls Unk. | Adjusted OR (95% CI) | P-value |
|---|---|---|---|---|---|---|
| Eat poultry liver pâté (not chicken) prepared outside home | 7 (1.5) | 0 | 2 (0.4) | 0 | 3.90 (0.79-19.16) | 0.094 |
| Frozen poultry (not chicken) purchased which was then prepared at home | 10 (2.2) | 38 | 16 (3.2) | 22 | 0.73 (0.32-1.68) | 0.464 |
| Fresh raw poultry (not chicken) purchased which was then prepared at home | 20 (4.4) | 51 | 28 (5.6) | 39 | 0.94 (0.51-1.73) | 0.847 |
| Raw poultry (not chicken) washed before preparation | 6 (1.3) | 66 | 7 (1.4) | 64 | 1.31 (0.41-4.21) | 0.645 |
| Raw poultry (not chicken) cut up in the kitchen | 10 (2.2) | 66 | 13 (2.6) | 64 | 1.05 (0.44-2.5) | 0.920 |
| Raw poultry (not chicken) handled in the kitchen | 9 (2.0) | 66 | 20 (4.0) | 65 | 0.62 (0.27-1.40) | 0.250 |
| Raw poultry (not chicken) at home - oven-cooked, roasted or grilled | 13 (2.9) | 66 | 19 (3.8) | 64 | 0.92 (0.44-1.94) | 0.830 |
| Poultry (not chicken) outside the home - oven-cooked, roasted or grilled | 25 (5.5) | 0 | 8 (1.6) | 0 | 4.20 (1.78-9.90) | 0.001 |
| Raw poultry (not chicken) at home - BBQ | 2 (0.4) | 66 | 0 (0) | 64 | ∞ | |
| Poultry (not chicken) outside the home - BBQ | 5 (1.1) | 0 | 2 (0.4) | 0 | 2.79 (0.54-14.43) | 0.256 |
| Raw poultry (not chicken) at home - stir fried | 3 (0.7) | 66 | 4 (0.8) | 64 | 1.16 (0.25-5.42) | 0.846 |

| (continued) | Cases, n (%) | Cases Unk. | Controls, n (%) | Controls Unk. | Adjusted OR (95% CI) | P-value |
|---|---|---|---|---|---|---|
| Poultry (not chicken) outside the home - stir fried | 4 (0.9) | 0 | 2 (0.4) | 0 | 2.67 (0.48-14.84) | 0.262 |
| Raw poultry (not chicken) at home - microwaved | 0 (0) | 66 | 1 (0.2) | 64 | 0 | |
| Poultry (not chicken) outside the home - microwaved | 2 (0.4) | 0 | 0 (0) | 0 | ∞ | |
| Raw poultry (not chicken) at home - stewed, slow cooked or steamed | 2 (0.4) | 66 | 3 (0.6) | 64 | 0.67 (0.11-4.19) | 0.665 |
| Poultry (not chicken) outside the home - stewed, slow cooked or steamed | 4 (0.9) | 0 | 2 (0.4) | 0 | 1.94 (0.32-11.90) | 0.473 |
| Raw poultry (not chicken) at home - deep fried | 0 (0) | 0 | 0 (0) | 64 | nd | nd |
| Poultry (not chicken) outside the home - deep fried | 5 (1.1) | 0 | 3 (0.6) | 0 | 2.24 (0.52-9.62) | 0.280 |
| Poultry (not chicken) lightly cooked (i.e. pinkish in the middle) | 2 (0.4) | 66 | 5 (1.0) | 66 | 0.58 (0.11-3.11) | 0.527 |
| **Foods eaten - other** | | | | | | |
| Eat either beef, pork, lamb, deer or rabbit | 318 (70.4) | 30 | 394 (78.8) | 23 | 0.59 (0.42-0.84) | 0.003 |
| Eat beef | 270 (59.7) | 30 | 332 (66.4) | 23 | 0.73 (0.54-0.98) | 0.039 |
| Eat beef undercooked (i.e. pinkish in the middle) | 52 (11.5) | 0 | 60 (12.0) | 22 | 1.15 (0.76-1.74) | 0.495 |

| (continued) | Cases, n (%) | Cases Unk. | Controls, n (%) | Controls Unk. | Adjusted OR (95% CI) | P-value |
|---|---|---|---|---|---|---|
| Eat beef oven-cooked, roasted or grilled | 150 (33.2) | 30 | 213 (42.6) | 23 | 0.64 (0.48-0.85) | 0.002 |
| Eat beef BBQ | 24 (5.3) | 30 | 18 (3.6) | 23 | 1.37 (0.71-2.61) | 0.345 |
| Eat beef stir fried | 30 (6.6) | 30 | 30 (6.0) | 23 | 1.21 (0.70-2.10) | 0.490 |
| Eat beef microwaved | 6 (1.3) | 30 | 12 (2.4) | 23 | 0.50 (0.17-1.46) | 0.206 |
| Eat beef deep fried | 3 (0.7) | 30 | 1 (0.2) | 23 | 2.93 (0.3-28.76) | 0.356 |
| Eat beef stewed, slow cooked or steamed | 125 (27.7) | 30 | 145 (29.0) | 23 | 1.00 (0.74-1.36) | 0.986 |
| Eat pork | 128 (28.3) | 30 | 187 (37.4) | 23 | 0.67 (0.50-0.90) | 0.007 |
| Eat pork undercooked (i.e. pinkish in the middle) | 8 (1.8) | 0 | 5 (1.0) | 22 | 1.46 (0.46-4.7) | 0.522 |
| Eat pork oven-cooked, roasted or grilled | 82 (18.1) | 30 | 136 (27.2) | 23 | 0.59 (0.43-0.82) | 0.002 |
| Eat pork BBQ | 11 (2.4) | 30 | 7 (1.4) | 23 | 1.58 (0.59-4.25) | 0.362 |
| Eat pork stir fried | 15 (3.3) | 30 | 28 (5.6) | 23 | 0.56 (0.29-1.10) | 0.095 |
| Eat pork microwaved | 3 (0.7) | 30 | 5 (1.0) | 23 | 0.73 (0.17-3.13) | 0.676 |
| Eat pork deep fried | 5 (1.1) | 30 | 2 (0.4) | 23 | 2.95 (0.56-15.44) | 0.200 |
| Eat pork stewed, slow cooked or steamed | 22 (4.9) | 30 | 35 (7.0) | 23 | 0.73 (0.41-1.3) | 0.282 |
| Eat lamb | 60 (13.3) | 30 | 77 (15.4) | 23 | 0.76 (0.51-1.11) | 0.158 |

| (continued) | **Cases, n (%)** | **Cases Unk.** | **Controls, n (%)** | **Controls Unk.** | **Adjusted OR (95% CI)** | **P-value** |
|---|---|---|---|---|---|---|
| Eat lamb undercooked (i.e. pinkish in the middle) | 8 (1.7) | 0 | 13 (2.6) | 22 | 0.50 (0.2-1.29) | 0.153 |
| Eat lamb oven-cooked, roasted or grilled | 44 (9.73) | 30 | 56 (11.2) | 23 | 0.82 (0.53-1.28) | 0.385 |
| Eat lamb BBQ | 1 (0.2) | 30 | 4 (0.8) | 23 | 0.21 (0.02-1.92) | 0.167 |
| Eat lamb stir fried | 2 (0.4) | 30 | 2 (0.4) | 23 | 1.18 (0.16-8.74) | 0.874 |
| Eat lamb microwaved | 2 (0.4) | 30 | 0 (0) | 23 | ∞ | |
| Eat lamb deep fried | 0 (0) | 30 | 0 (0) | 23 | nd | nd |
| Eat lamb stewed, slow cooked or steamed | 18 (4.0) | 30 | 19 (3.8) | 23 | 0.88 (0.43-1.78) | 0.718 |
| Eat deer or rabbit | 8 (1.8) | 30 | 11 (2.2) | 23 | 0.92 (0.36-2.36) | 0.860 |
| Eat deer or rabbit undercooked (i.e. Pinkish in the middle) | 3 (0.7) | 0 | 2 (0.4) | 23 | 0.90 (0.12-6.57) | 0.918 |
| Eat deer or rabbit oven-cooked, roasted or grilled | 3 (0.7) | 30 | 7 (1.4) | 22 | 0.44 (0.11-1.77) | 0.250 |
| Eat deer or rabbit BBQ | 1 (0.2) | 30 | 0 (0) | 23 | ∞ | |
| Eat deer or rabbit stir fried | 0 (0) | 30 | 1 (0.2) | 23 | 0 | |
| Eat deer or rabbit microwaved | 0 (0) | 30 | 0 (0) | 23 | nd | nd |
| Eat deer or rabbit deep fried | 0 (0) | 30 | 1 (0.2) | 23 | 0 | |

| (continued) | **Cases, n (%)** | **Cases Unk.** | **Controls, n (%)** | **Controls Unk.** | **Adjusted OR (95% CI)** | **P-value** |
|---|---|---|---|---|---|---|
| Eat deer or rabbit stewed, slow cooked or steamed | 4 (0.9) | 30 | 2 (0.4) | 23 | 3.10 (0.54-17.88) | 0.205 |
| Eat raw or lightly cooked fish / shell fish / sea food (e.g. fish, crab, prawns, mussels, oysters, calamari, sushi etc.) | 84 (18.6) | 27 | 104 (20.8) | 20 | 0.95 (0.67-1.33) | 0.753 |
| Eat any unpasteurised dairy products (incl. milk and cheese) | 54 (11.9) | 38 | 71 (14.2) | 39 | 0.86 (0.57-1.28) | 0.453 |

[a] Cases and controls unknown –the numbers of respondents that have not entered an answer to that particular question.

**Table 7.3 Domestic Multivariate Analysis of Univariate Risk Factors (P<0.05) No Weights**

| Factor | OR (95% CI) | P-value | PAF% (95% CI) |
|---|---|---|---|
| Ethnicity | 3.56 (1.03 - 12.23) | 0.044 | 71 (3 - 91) |
| Car <5 years old | 1.49 (1.11 - 2.00) | 0.008 | 17 (5 - 25) |
| Long term bowel condition | 1.58 (1.01 - 2.47) | 0.045 | 6 (0 - 10) |
| PPIs | 2.93 (1.94 -4.44) | <0.001 | 15 (11 - 18) |
| Contact with animals - overall | 0.73 (0.54 -0.98) | 0.036 | |
| Indoor swimming pool / toddler pool | 0.43 (0.24 -0.76) | 0.003 | |
| Eat chicken liver pâté prepared from raw at home | 3.41 (1.49 - 7.81) | 0.004 | 5 (2 - 6) |
| Raw chicken handled in the kitchen | 0.58 (0.41 - 0.82) | 0.002 | |
| Chicken lightly cooked (i.e. pinkish in the middle) | 6.48 (1.90 - 22.2) | 0.003 | 3 (1 - 3) |
| Chicken outside the home - Elsewhere | 1.69 (1.14 - 2.51) | 0.009 | 8 (2 - 11) |
| Poultry (other than chicken) outside the home - Restaurant | 5.80 (2.02 - 16.58) | 0.001 | 5 (3 - 5) |
| Eat pork oven-cooked, roasted or grilled | 0.59 (0.37 -0.95) | 0.028 | |

**Table 7.4 Domestic Multivariate Analysis of Univariate Risk Factors (P<0.05) Weights**

| Factor | OR (95% CI) | P-value | PAF% (95% CI) |
|---|---|---|---|
| Ethnicity | 4.60 (1.40 - 15.16) | 0.012 | 77 (28 - 92) |
| Car <5 years old | 1.73 (1.28 - 2.35) | <0.001 | 21 (11 - 29) |
| PPIs | 1.71 (1.28 - 2.30) | <0.001 | 10 (5 - 13) |
| Antiflatuents | 4.07 (1.30 - 12.7) | 0.016 | 2 (1 - 2) |
| Contact with Cats | 0.57 (0.39 -0.82) | 0.003 | |
| Indoor swimming pool / toddler pool | 0.42 (0.24 - 0.74) | 0.003 | |
| Eat chicken liver pâté prepared from raw at home | 3.69 (1.55 - 8.77) | 0.003 | 5 (2 - 6) |
| Raw chicken handled in the kitchen | 0.58 (0.41 - 0.81) | 0.001 | |
| Chicken lightly cooked (i.e. pinkish in the middle) | 4.88 (1.67 - 14.30) | 0.004 | 2 (1 - 3) |
| Chicken outside the home - Elsewhere | 1.74 (1.18 - 2.58) | 0.006 | 8 (3 - 12) |
| Poultry (other than chicken) outside the home - Restaurant | 6.66 (2.19 - 20.2) | 0.001 | 5 (3 - 5) |
| Eat beef oven-cooked, roasted or grilled | 0.62 (0.42 - 0.94) | 0.023 | |

**Table 7.5 Multivariate Model assignment**

| Dataset | Model | Percentage Correctly Assigned[a] |
|---|---|---|
| Domestic Case-control | Multivariate model no weights (P<0.157) | 68.7 |
| | Multivariate model no weights (P<0.05) | 67.8 |
| | Multivariate model with weights (P<0.157) | 68.5 |
| | Multivariate model with weights (P<0.05) | 67.2 |
| Foreign Travel Case-control | Multivariate model no weights (P<0.157) | 81.1 |
| | Multivariate model no weights (P<0.05) | 78.2 |
| | Multivariate model with weights (P<0.157) | 81.2 |
| | Multivariate model with weights (P<0.05) | 74.2 |

[a] This is the percentage correctly assigned for the 100 imputations that were carried out in SPSS.

## 7.2.4 Discussion – Domestic Case Control

**General details/Personal characteristics**

There was only one factor that came out of the analysis that can be associated with the socioeconomic status of the respondents – having a car < 5 years old. This was a statistically significant risk factor for campylobacteriosis in the univariate and multivariate (weighted and unweighted) analysis.

Individuals stating white ethnicity were more common among cases (98.9%) than controls (97.8%) but this was not statistically significant in the univariate analysis (Table 7.2). However, this became significant in all of the multivariate models (e.g. OR = 4.60 P=0.012) (Table 7.4). A previous study in England and Wales has shown that the Pakistani population has a higher incidence of campylobacteriosis than the white population but that the Indian and African populations have the lowest incidence (Gillespie, O'Brien et al. 2008). It is not possible to report the ethnicity of the non-white cases and controls in the current study because of the small numbers.

### Historical Health Conditions and Treatment

A pre-existing long term bowel (tummy) condition was identified as a risk factor in the univariate analysis (OR = 1.89, P=0.002). This was reported for a significant proportion of the population who returned questionnaires (17.3% cases and 9.4% controls). As mentioned above this was found to be a significant risk factor in the unweighted multivariate model (Table 7.3) but not the weighted one (Table 7.4).

The univariate analysis found that the usage of PPIs and the treatment of one or more of PPIs, H2 blockers, antacids or antiflatuents were significant risk factors. The univariate and all of the multivariate analysis models found that the use of PPIs was a risk factor. Indeed 23.5% of cases and 10.4% of controls were taking PPIs. Previous *Campylobacter* case-control studies from Scotland (Smith-Palmer, Cowden 2010), Denmark (Kuhn, Nielsen et al. 2018) the Netherlands (Doorduyn, Van den Brandhof et al. 2010) and Germany (Rosner, Schielke et al. 2017) have all identified PPI's as a risk factor. The population attributable fraction for this risk factor is 10% (5 – 13). This is a sizeable proportion of the population that acquire campylobacteriosis and may benefit from advice on how to avoid it. It is also worth noting that this is particularly important by age with only approximately 10% of cases being on a PPI aged under 25 years whereas approximately 25% of cases being >25 years (23% 25-64 years and 26% >64 years).

Although antiflatuents were not significantly associated with campylobacteriosis risk in the univariate analysis they were significant in the weighted multivariate analysis model (OR = 4.07 P=0.016) (Table 7.4). However, the population attributable fraction is small (1%) indicating that this is associated with only a very small number of cases.

The percentage of hospitalised cases in the domestic case-control study was 14.8%. The percentage of hospitalised cases for the historical health conditions and treatments were: antiflatuents 33.3%; long term bowel (tummy) condition 19.2%; PPIs 17.0%; H2 blockers 16.7% and antacids 7.4%. This suggests that there was an increased risk of hospitalisation with antiflatuents and a decreased risk for antacids. However, the number of hospitalised cases with antacids (n=4) and antiflatuents (n=4) is low and further data would be required to confirm this result.

### Travel History

Both travel (including an overnight stay) within Scotland and outside Scotland but within the UK were not significant factors for campylobacteriosis infection (P>0.05, Table 7.2). A previous study undertaken in Aberdeen City and Shire reported that an overnight stay out with the study area but within Scotland, England and Wales was also not a risk factor (Smith-Palmer, Cowden 2010). Whereas in Denmark visiting a weekend cottage was not associated with campylobacteriosis (Kuhn, Nielsen et al. 2018).

### Contact with animals

There were some results for contact with animals that were significant but this was not consistent across the univariate analysis and all of the multivariate

models. In the univariate analysis contact with animals overall (OR = 0.75 P=0.044) and the subset other animals (OR = 0.47 P=0.019) were both "protective". Contact with animals overall remained significant in the multivariate unweighted regression (Table 7.3) but not in the multivariate weighted models. In the multivariate weighted models contact with cats was found to be significantly protective (OR = 0.57 P=0.003) (Table 7.4).

A meta-analysis of campylobacteriosis case-control studies prior to 2004 found that of the 38 studies analysed direct contact with farm animals and pets were both risk factors (Domingues, Pires et al. 2012). A study from the Netherlands found that ownership of cats was a risk factor (Doorduyn, Van den Brandhof et al. 2010) whilst in Spain animal contact was a risk factor with a PAF of 19% but exposure to dogs/cats at home was not a risk factor (Fajo-Pascual, Godoy et al. 2010). Further, in an all-Ireland study contact with sheep was associated with campylobacteriosis risk (Danis, Di Renzi et al. 2009). The Aberdeen City and Shire study found that having a pet animal at home or an ill pet at home were both risk factors whilst farm animal contact was a risk factor in the univariate analysis but not so when there was adjustments (Smith-Palmer, Cowden 2010).

The message from the above case-control studies is that generally contact with animals, particularly farm animals that are known to shed *Campylobacter* asymptomatically (Ogden, Dallas et al. 2009), are a risk factor however the evidence for pet contact seems more indeterminate. The findings here that the animal contact risk factors that were significant were protective is at odds with the Aberdeen City and Shire study but this may be due to the different populations being considered (i.e. most of the population were from the central belt of Scotland and no children <5 years of age that are known to have a high incidence of campylobacteriosis were included in the study (N. J. Strachan, Gormley et al. 2009)).

**Exposure to water**

Overall water activity was found to be "protective" in the univariate analysis (OR = 0.55 P=0.006) but not in any of the multivariate analysis. Whereas indoor swimming pool/toddler pool was "protective" in the univariate analysis (OR = 0.44 P=0.002) (Table 7.2) and all of the multivariate analysis (e.g. OR = 0.42 P=0.003)(Table 7.4).

The Aberdeen City and Shire study found that being on a private as opposed to a public water supply was a risk factor and that the risk was greatest for children (Smith-Palmer, Cowden 2010). The current study did not find any of the water sources as a risk factor for campylobacteriosis. However, it should be noted that only approximately 5% of cases and controls were not on a public supply and hence the study may have lacked sufficient statistical power to detect a difference and also young children were not included. Other studies have also found non-public mains water not to be a risk factor. For example, in Ireland well water (Danis, Di Renzi et al. 2009) and in Spain all of tap, bottled or untreated water (Fajo-Pascual, Godoy et al. 2010) were all found not to be risk factors.

The Aberdeen City and Shire study did not find any water activity as a risk factor except for diving in the sea (Smith-Palmer, Cowden 2010) which had very few

cases reporting the exposure. In Ireland swimming or water sports in the sea (Danis, Di Renzi et al. 2009) as well as a meta-analysis of 38 studies across the world looking at recreational waters (Domingues, Pires et al. 2012) reported these as not being risk factors. Whilst in Denmark bathing in fresh water as well as children 1-5 years bathing in a paddling pool were risk factors (Kuhn, Nielsen et al. 2018).

It is unclear why in the current study that indoor swimming pool /toddler pool is "protective". Such pools are likely to be chlorinated, there is very little person to person transmission of *Campylobacter* and there are likely to be low risk of contamination from birds and other environmental vectors. As such it can be suggested that this type of activity can be low risk. However, for it to be protective is difficult to rationalise unless it is replacing an activity that is of higher risk or alternately it could be a proxy for other healthy behaviours.

**Foods eaten - chicken**

The univariate analysis indicates that eating chicken prepared at home is "protective" (OR = 0.70 P=0.016) whilst eating chicken outside the home is neither protective or a risk (OR=0.05 P>0.05). These results immediately suggest that chicken is a complex risk factor.

The previous Aberdeen City and Shire study had found eating chicken outside the home also to be a risk factor but eating chicken at home was not (Smith-Palmer, Cowden 2010). The international meta-analysis study found that eating chicken out at a restaurant was a risk factor (Domingues, Pires et al. 2012).

A number of studies found that eating chicken (either at home or out) was a risk factor. For example in Denmark eating whole, boneless fillets or chicken thighs (Kuhn, Nielsen et al. 2018), in Ireland and the Netherlands eating any chicken (Danis, Di Renzi et al. 2009, Doorduyn, Van den Brandhof et al. 2010)) were risk factors, whilst the meta-analysis of 38 case-control studies did not (Domingues, Pires et al. 2012). Although not provided in Table 7.2 eating any chicken is not a risk factor in the current study (P>0.05).

Considering consumption of chicken at home. Eating chicken liver pâté prepared from raw at home is a risk factor in the univariate analysis (OR = 3.82 P<0.001) (Table 7.2) as well as all of the multivariate models (e.g. OR = 3.69 P=0.003) with a population attributable fraction of 0.05 (Table 7.4). Chicken liver pâté is well established as being a potential risk of campylobacteriosis because many recipes indicate it should be consumed lightly cooked (Jones, Rigby et al. 2016). There have been a number of outbreaks associated with it in the UK (Little, Gormley et al. 2010, Forbes, Gormley et al. 2009) and it is known that the types of *Campylobacter* found in chicken livers are also the types commonly found in human disease (N. J. C. Strachan, MacRae et al. 2012). However eating chicken liver pâté prepared outside the home was not found to be a risk factor and it is worth noting the low percentage of individuals (1.3% of cases and 1% of controls) exposed. It is possible that restaurants etc. are more aware of the risks and cook the livers more thoroughly or that the low level of exposure may have insufficient statistical power.

Raw chicken handled in the kitchen was "protective" in the univariate (OR = 0.71, P<0.023, Table 7.2) and all of the multivariate models (e.g. OR = 0.58

P=0.001, Table 7.3) except for the unweighted, P<0.157 model where the p-value was 0.06. It is difficult to conceptualise why this would be protective unless it was replacing a higher risk activity (e.g. eating out) or if handling conferred immunity or if hygiene in the domestic kitchen was very good. Previous studies had found that handling raw chicken at home was not a risk factor (e.g. in Denmark (Kuhn, Nielsen et al. 2018)).

Eating chicken lightly cooked, pinkish in the middle was a risk factor in the univariate (OR = 4.43, P=0.017, Table 7.2) and all of the multivariate models (e.g. OR = 4.88, P=0.006, Table 7.3) with a PAF of 0.02 (Table 7.4). A number of other studies had also found this to be a risk factor (e.g. meta-analysis of 38 case-control studies (Domingues, Pires et al. 2012), in the Netherlands (Doorduyn, Van den Brandhof et al. 2010) and Ireland (Danis, Di Renzi et al. 2009)). But some others had not for example in Denmark (Kuhn, Nielsen et al. 2018) and Spain (Fajo-Pascual, Godoy et al. 2010). However, it would seem to be pertinent to ensure that the population are aware of the potential risk of undercooked chicken but this appears only to be an issue for a small proportion of the population.

It appears that the apparent additional risk of eating chicken outside the home is not associated with takeaway or fast food outlets or visiting restaurants but is elsewhere (Table 7.2). Indeed eating chicken outside the home (elsewhere comprised predominantly from a friend or relatives house or a community/family gathering) is significant in the univariate and all of the multivariate models. This is in agreement with the all-Ireland study (Danis, Di Renzi et al. 2009).

Eating chicken consumed at a barbecue has previously been associated with campylobacteriosis (Doorduyn, Van den Brandhof et al. 2010). In the current study eating chicken prepared from raw at a home BBQ was a risk factor in the univariate analysis (OR = 3.16 P<0.019, Table 7.2) but was not significant in any of the multivariate models. Further eating chicken at a BBQ away from home was not a risk factor.

Not washing raw chicken has been a prominent food safety message but was not a risk factor despite 61 cases and 65 controls stating that they carried out this practice.

**Foods eaten - poultry other than chicken**

Eating poultry other than chicken outside the home was found to be a risk factor in the univariate analysis (OR = 2.5, P=0.004, Table 7.2) as was eating poultry at a restaurant and eating poultry elsewhere. However, eating poultry (other than chicken) at a restaurant was also a significant risk factor in all of the multivariate analysis (e.g. OR=6.66, P=0.001, Table 7.4). Unfortunately, the species of the poultry was not requested, though in Scotland it is most likely that the main species consumed are turkey and duck whilst it is more unusual to eat game birds such as quail, partridge, pheasant etc. In 2018 UK poultry meat production was 1.9M tonne comprising broilers (86%), boiling fowl (4.0%), turkeys (8.3%) and ducks (1.6%) (DEFRA 2019). In the univariate analysis if it was oven cooked, roasted or grilled it was a risk factor but this was no longer significant in any of the multivariate models. Although not presented in Table 7.2 eating any poultry (other than chicken) was not a risk factor (P>0.05).

Previous studies reporting consumption of particular species of poultry other than chicken did not break this down where the food was eaten. However eating any turkey was found to be protective in an Irish (Danis, Di Renzi et al. 2009) and a Danish study (Kuhn, Nielsen et al. 2018). In a meta-analysis of 38 studies, eating any poultry or eating any poultry at home was not a risk factor (Domingues, Pires et al. 2012). Further, consumption of duck was not found to be a risk factor in an Irish study (Danis, Di Renzi et al. 2009). Further investigation of why consumption of poultry (other than chicken) at a restaurant is a risk factor is worth pursuing particularly because the PAF is 5%.

It is worth noting that eating poultry other than chicken lightly cooked (i.e. pinkish in middle) is not a risk factor, as is consumption of liver pâté, which contrasts to what is found from chicken. It may be that since these are rare exposures in the population the numbers in the study do not have sufficient statistical power to detect the risk.

**Foods Eaten - Other**

A number of these foods were significant in the univariate analysis but all were protective (e.g. ate either beef, pork, lamb or deer OR = 0.59 P=0.003, or ate beef or ate pork) (Table 7.2). However in the multivariate analysis only ate beef in the weighted P<0.05 model and ate pork in the unweighted P=0.05 model were significant. It is possible that the reason that these are "protective" is because if they are being consumed then it means that the individuals are not eating foods of higher risk. It is known that for beef the prevalence of *Campylobacter* is low whilst for pork, although *C. coli* is shed by pigs, many of the sequence types present do not appear to be found in sick humans (N. J. C. Strachan, Rotariu et al. 2013).

**Vegetarian**

In the study there were 15 cases and 24 controls that were vegetarian. Although the OR was <1.0 it was not significant (P>0.05).

## 7.2.5 Results – Foreign Travel Case Control

### 7.2.5.1 Foreign travel associated univariate and multivariate logistic regression

Overall there were 146 cases (24.4%) and 52 controls (9.4%) that travelled abroad with an overnight stay (Fig. 6.1). This this was a significant risk for human campylobacteriosis (OR = 4.1, 95% CI = 2.9 to 5.8, P<0.001). A number of previous studies have also found foreign travel as a risk factor: the Aberdeen City and Shire study (Smith-Palmer, Cowden 2010); a recent Danish case-control study (Kuhn, Nielsen et al. 2018) and a Dutch study (Doorduyn, Van den Brandhof et al. 2010).

The results of the univariate analysis for those variables selected as possible adjustments for confounding are presented in Table 7.1. Both season and Scottish index of multiple deprivation (SIMD) were not significant in the

univariate analysis. However, risk of campylobacteriosis increased significantly (P = 0.042) when travelling abroad if aged between 25-64 compared with the reference group aged 65+ (Table 7.1). It was also found that males were less likely (P=0.031) than females to contract campylobacteriosis when travelling abroad (Table 7.1). Hence, both age and sex were included as adjustments in the multivariate analysis.

The univariate analysis results are presented in Table 7.6. In total 2 factors were a risk whilst 8 were "protective". The multivariate analysis for the P<0.05 model unweighted and weighted are presented in Tables 7.7 and 7.8. The corresponding tables for the P<0.157 weighted and unweighted multivariate models are given in the Annex (Tables A7.3 and A7.4).

**Table 7.6 Foreign Travel Univariate Analysis**

| General Details - Personal Characteristics | Cases n (%) | Cases Unk. | Controls, n (%) | Controls Unk. | Adjusted OR (95%)[a,b] | P-value |
|---|---|---|---|---|---|---|
| N | 146 | | 52 | | | |
| **Season** | | | | | | |
| Summer | 62 (42.5) | 0 | 25 (48.1) | 0 | 0.6 (0.3-1.19) | 0.146 |
| Rest of year (Reference) | 84 (57.5) | 0 | 27 (51.9) | 0 | | |
| **Age (years)** | | | | | | |
| 5-14 | 3 (2.1) | 3 | 2 (3.8) | 0 | 1.17 (0.17-8.04) | 0.874 |
| 15-24 | 9 (6.2) | 3 | 3 (5.8) | 0 | 1.74 (0.40-7.64) | 0.462 |
| 25-64 | 108 (74.0) | 3 | 32 (61.5) | 0 | 2.05 (0.95-4.43) | 0.069 |
| 65+ (Reference) | 23 (15.8) | 3 | 15 (28.8) | 0 | | |
| **Sex** | | | | | | |
| Male | 56 (38.4) | 0 | 29 (55.8) | 0 | 0.52 (0.27-0.99) | 0.048 |
| Female (Reference) | 90 (61.6) | 0 | 23 (44.2) | 0 | | |
| **Ethnicity** | | | | | | |
| White | 143 (97.9) | 0 | 50 (96.1) | 0 | 2.02 (0.32-12.80) | 0.456 |
| Other (Reference) | 3 (2.0) | 0 | 2 (3.8) | 0 | | |

| (continued) | **Cases n (%)** | **Cases Unk.** | **Controls, n (%)** | **Controls Unk.** | **Adjusted OR (95%)[a,b]** | **P-value** |
|---|---|---|---|---|---|---|
| **SIMD** | | | | | | |
| 5 | 123 (84.2) | 1 | 45 (86.5) | 2 | 0.63 (0.22-1.82) | 0.397 |
| 1 (Reference) | 22 (15.1) | 1 | 5 (9.6) | 2 | | |
| **Rurality, benefits and income** | | | | | | |
| R | 11 (7.5) | 2 | 5 (9.6) | 2 | nd | |
| pU | 31 (21.2) | 2 | 16 (30.7) | 2 | nd | |
| U | 102 (69.9) | 2 | 29 (55.8) | 2 | nd | |
| Benefits | 11 (7.5) | 0 | 4 (7.7) | 0 | 0.89 (0.27-3.01) | 0.857 |
| Household income >£47k/year | 69 (47.3) | 5 | 31 (59.6) | 1 | 0.49 (0.23-1.01) | 0.054 |
| **Occupation** | | | | | | |
| Retired | 44 (30.1) | 0 | 16 (30.8) | 0 | 2.01 (0.73-5.56) | 0.178 |
| School aged | 3 (2.1) | 0 | 1 (1.9) | 0 | ∞ | |
| Student | 5 (3.4) | 0 | 2 (3.8) | 0 | 0.59 (0.08-4.34) | 0.605 |

| (continued) | **Cases n (%)** | **Cases Unk.** | **Controls, n (%)** | **Controls Unk.** | **Adjusted OR (95%)[a,b]** | **P-value** |
|---|---|---|---|---|---|---|
| Unemployed | 2 (1.4) | 0 | 1 (1.9) | 0 | 0.43 (0.04-5.01) | 0.501 |
| Professional job | 75 (51.4) | 0 | 28 (53.8) | 0 | 0.56 (0.25-1.27) | 0.165 |
| Car <5 years old | 75 (51.4) | 0 | 29 (55.8) | 0 | 0.84 (0.43-1.64) | 0.837 |
| House >= 3 bedrooms | 101 (69.2) | 0 | 40 (76.9) | 0 | 0.72 (0.34-1.56) | 0.410 |
| Sit/Sat on a committee or council | 23 (15.8) | 0 | 9 (17.3) | 0 | 0.97 (0.40-2.34) | 0.949 |
| None of the above | 9 (6.2) | 0 | 2 (3.8) | 0 | 1.6 (0.32-7.95) | 0.567 |
| **Historical health conditions and treatment** | | | | | | |
| Long term bowel condition | 28 (19.2) | 0 | 6 (11.5) | 0 | 1.64 (0.62-4.36) | 0.317 |
| Other medical condition | 48 (32.9) | 0 | 20 (38.5) | 0 | 0.95 (0.47-1.92) | 0.881 |
| PPIs | 31 (21.2) | 0 | 7 (13.5) | 0 | 1.86 (0.74-4.70) | 0.188 |
| H2-blockers | 2 (1.4) | 0 | 3 (5.8) | 0 | 0.2 (0.03-1.33) | 0.096 |
| Antacids | 21 (14.4) | 0 | 9 (17.3) | 0 | 0.72 (0.30-1.74) | 0.464 |
| Antiflatuents | 4 (2.7) | 0 | 0 (0) | 0 | ∞ | |
| PPIs, H2-blockers, Antacids, Antiflatuents | 47 (32.2) | 0 | 14 (26.9) | 0 | 1.22 (0.59-2.56) | 0.589 |
| Antibiotics | 14 (9.6) | 0 | 3 (5.8) | 0 | 2.15 (0.58-8.06) | 0.254 |
| Medicine (other) | 58 (39.7) | 0 | 25 (48.1) | 0 | 0.82 (0.41-1.62) | 0.697 |

| (continued) | Cases n (%) | Cases Unk. | Controls, n (%) | Controls Unk. | Adjusted OR (95%)[a,b] | P-value |
|---|---|---|---|---|---|---|
| **Travel** | | | | | | |
| Travel within Scotland | 14 (9.6) | 2 | 12 (23.1) | 1 | nd | |
| Travel outside Scotland | 146 (100) | 0 | 52 (100) | 0 | nd | |
| Foreign travel | 146 (100) | 0 | 52 (100) | 0 | nd | |
| **Contact with animals** | | | | | | |
| Contact with types of animals | 53 (36.3) | 0 | 25 (48.1) | 0 | 0.51 (0.26-1.00) | 0.051 |
| Dogs | 27 (18.5) | 0 | 21 (40.4) | 0 | 0.29 (0.14-0.61) | 0.001 |
| Cats | 19 (13.0) | 0 | 7 (13.5) | 0 | 0.83 (0.31-2.20) | 0.705 |
| Birds/Poultry | 5 (3.4) | 0 | 0 (0) | 0 | ∞ | |
| Farm animals(cattle, sheep, goats, horses, donkeys, pigs) | 6 (4.1) | 0 | 2 (3.8) | 0 | 0.54 (0.10-3.02) | 0.484 |
| Other animals | 8 (5.5) | 0 | 1 (1.9) | 0 | 1.95 (0.23-16.82) | 0.543 |
| Contact with ill animal | 1 (0.7) | 0 | 0 (0) | 0 | ∞ | |
| Touch animal faeces | 11 (7.5) | 9 | 4 (7.7) | 5 | 0.83 (0.25-2.82) | 0.768 |
| **Exposure to water** | | | | | | |
| Water activity | 48 (32.9) | 1 | 13 (25.0) | 0 | 1.28 (0.61-2.72) | 0.514 |
| Indoor swimming pool / toddler pool | 10 (6.8) | 1 | 4 (7.7) | 0 | 0.92 (0.27-3.18) | 0.898 |
| Outdoor swimming pool / paddling pool / theme park water ride / splash park | 34 (23.3) | 1 | 7 (13.5) | 0 | 1.78 (0.71-4.44) | 0.215 |

| (continued) | Cases n (%) | Cases Unk. | Controls, n (%) | Controls Unk. | Adjusted OR (95%)[a,b] | P-value |
|---|---|---|---|---|---|---|
| Loch/lake/pond/stream/river/burn (e.g. swimming, canoeing, diving, fishing) | 7 (4.8) | 1 | 1 (1.9) | 0 | 1.79 (0.20-15.67) | 0.599 |
| Sea (e.g. diving, sailing, surfing, jet ski, fishing) | 15 (10.3) | 1 | 4 (7.7) | 0 | 1.23 (0.37-4.12) | 0.735 |
| Other water activity | 4 (2.7) | 1 | 1 (1.9) | 0 | 1.22 (0.13-11.54) | 0.861 |
| Water source (public mains) | 138 (94.5) | 1 | 51 (98.1) | 0 | 0.42 (0.05-3.52) | 0.420 |
| Water source (private-spring) | 0 (0) | 1 | 0 (0) | 0 | | |
| Water source (private-well) | 1 (0.7) | 1 | 2 (3.8) | 0 | 0.2 (0.02-2.49) | 0.212 |
| Water source (River/stream/lake/loch/pond/melted snow (not boiled) | 2 (1.4) | 1 | 1 (1.9) | 0 | 0.56 (0.05-6.59) | 0.645 |
| **Exposure to food** | | | | | | |
| **Foods eaten - chicken** | | | | | | |
| Eat chicken prepared at home | 43 (29.5) | 9 | 28 (53.8) | 1 | 0.40 (0.20-0.80) | 0.009 |
| Eat chicken outside the home | 92 (63.0) | 0 | 23 (44.2) | 0 | 2.03 (1.02-4.03) | 0.043 |
| Chicken outside the home - Restaurant | 73 (50.0) | 0 | 17 (32.7) | 0 | 1.95 (0.97-3.90) | 0.060 |
| Chicken outside the home - Take away or Fast food | 19 (13.0) | 0 | 6 (11.5) | 0 | 1.04 (0.37-2.89) | 0.939 |
| Chicken outside the home -Elsewhere | 44 (30.1) | 0 | 8 (15.4) | 0 | 2.16 (0.92-5.07) | 0.077 |
| Eat chicken liver pâté prepared from raw at home | 4 (2.7) | 0 | 1 (1.9) | 0 | 1.53 (0.16-14.6) | 0.711 |
| Eat chicken liver pâté prepared outside home | 3 (2.1) | 0 | 1 (1.9) | 0 | 0.80 (0.08-8.12) | 0.847 |

| (continued) | Cases, n (%) | Cases Unk. | Controls, n (%) | Controls Unk. | Adjusted OR (95%)[a,b] | P-value |
|---|---|---|---|---|---|---|
| Frozen chicken purchased which was then prepared at home | 9 (6.16) | 18 | 6 (11.5) | 5 | 0.68 (0.22-2.14) | 0.512 |
| Fresh raw chicken purchased which was then prepared at home | 42 (28.8) | 11 | 24 (46.2) | 1 | 0.49 (0.25-0.98) | 0.044 |
| Raw chicken washed before preparation | 8 (5.5) | 15 | 5 (9.6) | 3 | 0.78 (0.23-2.67) | 0.697 |
| Raw chicken cut up in the kitchen | 32 (21.9) | 16 | 19 (36.5) | 3 | 0.52 (0.25-1.08) | 0.079 |
| Raw chicken handled in the kitchen | 25 (17.1) | 15 | 17 (32.7) | 4 | 0.43 (0.20-0.93) | 0.031 |
| Raw chicken at home - oven-cooked, roasted or grilled | 21 (14.4) | 14 | 13 (25.0) | 3 | 0.57 (0.25-1.30) | 0.179 |
| Chicken outside the home - oven-cooked, roasted or grilled | 67 (45.9) | 0 | 18 (34.6) | 0 | 1.43 (0.72-2.84) | 0.301 |
| Raw chicken at home - BBQ | 2 (1.4) | 0 | 2 (3.8) | 3 | 0.20 (0.03-1.52) | 0.119 |
| Chicken outside the home - BBQ | 13 (8.9) | 0 | 0 (0) | 0 | ∞ | |
| Raw chicken at home - stir fried | 15 (10.3) | 14 | 9 (17.3) | 3 | 0.54 (0.21-1.39) | 0.202 |
| Chicken outside the home - stir fried | 19 (13.0) | 0 | 2 (3.8) | 0 | 2.92 (0.64-13.36) | 0.167 |
| Raw chicken at home - microwaved | 0 (0) | 14 | 0 (0) | 3 | ∞ | |
| Chicken outside the home - microwaved | 5 (3.4) | 0 | 0 (0) | 0 | ∞ | |
| Raw chicken at home - stewed, slow cooked or steamed | 2 (1.4) | 14 | 4 (7.7) | 3 | 0.14 (0.02-0.83) | 0.031 |
| Chicken outside the home - stewed, slow cooked or steamed | 14 (9.6) | 0 | 1 (1.9) | 0 | 5.66 (0.71-44.77) | 0.101 |

| (continued) | Cases n (%) | Cases Unk. | Controls, n (%) | Controls Unk. | Adjusted OR (95%)[a,b] | P-value |
|---|---|---|---|---|---|---|
| Raw chicken at home - deep fried | 1 (0.7) | 14 | 0 (0) | 3 | ∞ | |
| Chicken outside the home - deep fried | 12 (8.2) | 0 | 2 (3.8) | 0 | 3.62 (0.60-21.86) | 0.161 |
| Chicken lightly cooked (i.e. pinkish in the middle) | 0 (0) | 22 | 0 (0) | 3 | 0 | |
| **Foods eaten – poultry other than chicken** | | | | | | |
| Eat poultry other than chicken prepared at home | 3 (2.1) | 6 | 6 (11.5) | 1 | 0.09 (0.02-0.50) | 0.006 |
| Eat poultry other than chicken prepared outside the home | 20 (13.7) | 0 | 6 (11.5) | 0 | 1.15 (0.42-3.14) | 0.792 |
| Poultry (not chicken) outside the home - Restaurant | 18 (12.3) | 0 | 5 (9.6) | 0 | 1.28 (0.44-3.78) | 0.651 |
| Poultry (not chicken) outside the home - Take away or Fast food | 2 (1.4) | 0 | 1 (1.9) | 0 | 0.92 (0.07-11.40) | 0.950 |
| Poultry (not chicken) outside the home - Elsewhere | 8 (5.5) | 0 | 2 (3.8) | 0 | 1.21 (0.24-6.10) | 0.817 |
| Eat poultry liver pâté (not chicken) which was prepared from raw at home | 0 (0) | 7 | 0 (0) | 1 | ∞ | |
| Eat poultry liver pâté (not chicken) prepared outside home | 1 (0.7) | 0 | 0 (0) | 0 | 0 | |
| Frozen poultry (not chicken) purchased which was then prepared at home | 0 (0) | 7 | 2 (3.8) | 2 | 0 | |
| Fresh raw poultry (not chicken) purchased which was then prepared at home | 3 (2.1) | 9 | 6 (11.5) | 4 | 0 | |
| Raw poultry (not chicken) washed before preparation | 1 (0.7) | 16 | 1 (1.9) | 8 | 0.42 (0.03-7.15) | 0.551 |

| (continued) | **Cases n (%)** | **Cases Unk.** | **Controls, n (%)** | **Controls Unk.** | **Adjusted OR (95%)[a,b]** | **P-value** |
|---|---|---|---|---|---|---|
| Raw poultry (not chicken) cut up in the kitchen | 2 (1.4) | 16 | 4 (7.7) | 8 | 0.18 (0.03-1.11) | 0.064 |
| Raw poultry (not chicken) handled in the kitchen | 1 (0.7) | 16 | 4 (7.7) | 8 | 0.07 (0.01-0.71) | 0.024 |
| Raw poultry (not chicken) at home - oven-cooked, | 3 (2.1) | 16 | 4 (7.7) | 8 | 0.24 (0.04-1.46) | 0.121 |
| Poultry (not chicken) outside the home - oven- | 12 (8.2) | 0 | 4 (7.7) | 0 | 1.11 (0.33-3.73) | 0.865 |
| Raw poultry (not chicken) at home - BBQ | 1 (0.7) | 16 | 0 (0) | 8 | ∞ | |
| Poultry (not chicken) outside the home - BBQ | 3 (2.1) | 0 | 1 (1.9) | 0 | 1.12 (0.11-11.49) | 0.922 |
| Raw poultry (not chicken) at home - stir fried | 1 (0.7) | 16 | 2 (3.8) | 8 | 0.11 (0.01-1.31) | 0.081 |
| Poultry (not chicken) outside the home - stir fried | 3 (2.1) | 0 | 0 (0) | 0 | ∞ | |
| Raw poultry (not chicken) at home - microwaved | 0 (0) | 16 | 1 (1.9) | 8 | 0 | |
| Poultry (not chicken) outside the home - microwaved | 0 (0) | 0 | 1 (1.9) | 0 | 0 | |
| Raw poultry (not chicken) at home - stewed, slow cooked or steamed | 1 (0.7) | 16 | 1 (1.9) | 8 | 0.42 (0.03-7.15) | 0.551 |
| Poultry (not chicken) outside the home - stewed, slow cooked or steamed | 4 (2.7) | 0 | 2 (3.8) | 0 | 0.57 (0.09-3.56) | 0.544 |
| Raw poultry (not chicken) at home - deep fried | 1 (0.7) | 0 | 0 (0) | 8 | ∞ | |
| Poultry (not chicken) outside the home - deep fried | 3 (2.1) | 0 | 0 (0) | 0 | ∞ | |
| Poultry (not chicken) lightly cooked (i.e. pinkish in the middle) | 0 (0) | 16 | 1 (1.9) | 10 | 0 | |

| (continued) | **Cases n (%)** | **Cases Unk.** | **Controls, n (%)** | **Controls Unk.** | **Adjusted OR (95%)[a,b]** | **P-value** |
|---|---|---|---|---|---|---|
| **Foods eaten - other** | | | | | | |
| Eat either beef, pork, lamb, deer or rabbit | 105 (71.9) | 8 | 41 (78.8) | 1 | 0.94 (0.41-2.16) | 0.883 |
| Eat beef | 83 (56.8) | 8 | 37 (71.2) | 1 | 0.61 (0.3-1.26) | 0.183 |
| Eat beef undercooked (i.e. pinkish in the middle) | 31 (21.2) | 0 | 12 (23.1) | 0 | 0.86 (0.39-1.89) | 0.704 |
| Eat beef oven-cooked, roasted or grilled | 60 (41.1) | 8 | 31 (59.6) | 1 | 0.54 (0.27-1.06) | 0.075 |
| Eat beef BBQ | 11 (7.5) | 8 | 4 (7.7) | 1 | 0.98 (0.28-3.37) | 0.971 |
| Eat beef stir fried | 15 (10.3) | 8 | 1 (1.9) | 1 | 5.46 (0.70-42.85) | 0.106 |
| Eat beef microwaved | 1 (0.7) | 8 | 2 (3.8) | 1 | 0 | |
| Eat beef deep fried | 1 (0.7) | 8 | 0 (0) | 1 | ∞ | |
| Eat beef stewed, slow cooked or steamed | 20 (13.7) | 8 | 9 (17.3) | 1 | 0.81 (0.33-2.00) | 0.651 |
| Eat pork | 46 (31.5) | 8 | 23 (44.2) | 1 | 0.68 (0.34-1.35) | 0.271 |
| Eat pork undercooked (i.e. pinkish in the middle) | 1 (0.7) | 0 | 1 (1.9) | 0 | 0.43 (0.03-7.07) | 0.554 |
| Eat pork oven-cooked, roasted or grilled | 37 (25.3) | 8 | 17 (32.7) | 1 | 0.74 (0.36-1.53) | 0.417 |
| Eat pork BBQ | 4 (2.7) | 8 | 2 (3.8) | 1 | 0.65 (0.11-3.79) | 0.636 |
| Eat pork stir fried | 4 (2.7) | 8 | 4 (7.7) | 1 | 0.34 (0.07-1.64) | 0.179 |
| Eat pork microwaved | 1 (0.7) | 8 | 0 (0) | 1 | ∞ | |

| (continued) | Cases n (%) | Cases Unk. | Controls, n (%) | Controls Unk. | Adjusted OR (95%)[a,b] | P-value |
|---|---|---|---|---|---|---|
| Eat pork deep fried | 1 (0.7) | 8 | 1 (1.9) | 1 | 0.8 (0.04-15.48) | 0.880 |
| Eat pork stewed, slow cooked or steamed | 7 (4.8) | 8 | 2 (3.8) | 1 | 1.46 (0.29-7.41) | 0.646 |
| Eat lamb | 33 (22.6) | 8 | 6 (11.5) | 1 | 2.62 (1.01-6.85) | 0.049 |
| Eat lamb undercooked (i.e. pinkish in the middle) | 2 (1.4) | 0 | 2 (3.8) | 0 | 0.58 (0.07-4.51) | 0.599 |
| Eat lamb oven-cooked, roasted or grilled | 25 (17.1) | 8 | 4 (7.7) | 1 | 3.02 (0.97-9.44) | 0.058 |
| Eat lamb BBQ | 3 (2.1) | 8 | 0 (0) | 1 | ∞ | |
| Eat lamb stir fried | 2 (1.4) | 8 | 0 (0) | 1 | ∞ | |
| Eat lamb microwaved | 0 (0) | 8 | 0 (0) | 1 | ∞ | |
| Eat lamb deep fried | 1 (0.7) | 8 | 0 (0) | 1 | ∞ | |
| Eat lamb stewed, slow cooked or steamed | 6 (4.1) | 8 | 2 (3.8) | 1 | 1.22 (0.23-6.50) | 0.817 |
| Eat deer or rabbit | 2 (1.4) | 8 | 1 (1.9) | 1 | 1.25 (0.1-16.05) | 0.863 |
| Eat deer or rabbit undercooked (i.e. pinkish in the middle) | 0 (0) | 0 | 1 (1.9) | 0 | 0 | |
| Eat deer or rabbit oven-cooked, roasted or grilled | 2 (1.4) | 8 | 1 (1.9) | 1 | ∞ | |
| Eat deer or rabbit BBQ | 0 (0) | 8 | 0 (0) | 1 | ∞ | |
| Eat deer or rabbit stir fried | 0 (0) | 8 | 0 (0) | 1 | ∞ | |
| Eat deer or rabbit microwaved | 0 (0) | 8 | 0 (0) | 1 | ∞ | |
| Eat deer or rabbit deep fried | 0 (0) | 8 | 0 (0) | 1 | ∞ | |
| Eat deer or rabbit stewed, slow cooked or steamed | 0 (0) | 8 | 0 (0) | 1 | ∞ | |

| (continued) | **Cases n (%)** | **Cases Unk.** | **Controls, n (%)** | **Controls Unk.** | **Adjusted OR (95%)[a,b]** | **P-value** |
|---|---|---|---|---|---|---|
| Eat raw or lightly cooked fish / shell fish / sea food (e.g. fish, crab, prawns, mussels, oysters, calamari, sushi etc.) | 48 (32.9) | 7 | 16 (30.8) | 1 | 1.13 (0.55-2.33) | 0.732 |
| Eat any unpasteurised dairy products (incl. milk and cheese) | 26 (17.8) | 8 | 5 (9.6) | 5 | 1.61 (0.57-4.60) | 0.372 |

a – adjusted odds ratio by age and sex,
b – on occasions the OR may unexpectedly be zero or infinity because the adjustment variable(s) in the relevant cases or controls may be unknown and therefore not included in the analysis.
nd – not done
∞ - infinity as divide by zero.

**Table 7.7 Foreign Travel Multivariate Analysis No Weights P<0.05**

| Factor | Adjusted OR (95% CI) | P-value |
|---|---|---|
| Household income >£47k/year | 0.45 (0.21 - 0.99) | 0.046 |
| H2 blockers | 0.09 (0.01 - 0.67) | 0.018 |
| Dogs contact | 0.32 (0.15 - 0.69) | 0.003 |
| Eat lamb | 3.11 (1.09 - 8.90) | 0.034 |

**Table 7.8 Foreign Travel Multivariate Analysis with Weights P<0.05**

| Factor | Adjusted OR (95% CI) | P-value |
|---|---|---|
| Eat chicken prepared at home | 0.36 (0.16 - 0.78) | 0.010 |

## 7.2.6 Discussion – Foreign Travel Case Control

**General details/Personal characteristics**

After adjustment the only variable that was significant in the univariate analysis was sex, with males having a lower risk of campylobacteriosis when travelling abroad (P=0.048) (Table 7.6). However, household income >£47k per year was found to be "protective" in the unweighted models but was not significant in the weighted models.

**Historical Health Conditions and Treatment -** None were significant in the univariate analysis but H2 blockers were "protective" in the multivariate unweighted models but not the weighted ones.

**Travel History -** Analysis of the travel variables were not conducted because this focusses on foreign travel cases only.

**Contact with animals**

Contact with a dog decreased the risks of campylobacteriosis in the univariate analysis (Table 7.6). This finding remained in the unweighted multivariate models (Table 7.7 and Table A7.3) but was not significant in the weighted multivariate models (Table 7.8 and Table A7.4). It is unclear why contact with a dog should decrease the risk. In a Dutch case-control study ownership of a dog was not significant (Doorduyn, Van den Brandhof et al. 2010). In the Aberdeen City and Shire study having a pet at home was a risk factor (OR = 1.23, p=0.02) (Smith-Palmer, Cowden 2010). It is unclear in the current study whether the dog contact was with a foreign dog, taking one's own dog abroad or being in contact with a dog in Scotland (i.e. during the 7 days before falling ill (cases) or being asked to complete the questionnaire (controls)).

**Exposure to water** - none were significant in the univariate analysis

**Foods Eaten - chicken**

In the univariate analysis (Table 7.6) eating chicken prepared at home was found to be "protective" as was: fresh raw chicken which was purchased and then prepared at home; raw chicken handled in the kitchen and raw chicken at home – stewed, slow cooked or steamed. Only eating chicken prepared at home remained significant in the weighted multivariate models (Table 7.8 and A7.4) but was not significant in the unweighted models.

Since foreign travel cases and controls are being considered it would be expected that the numbers associated with eating chicken at home would be small but they are not (43 cases and 28 controls). Hence, this indicates that part of the time prior to being infected (cases) or completing the questionnaire (controls) individuals were at home. It can be argued that since it is known travelling abroad is a risk factor, that if the trip is short or only encompasses only part of the incubation period then the likelihood of illness will be lower. If this was the case then it would be expected eating other foods at home as being protective as this would be a proxy for not being abroad. The same result was obtained for eating poultry other than chicken prepared at home. Unfortunately, however, this question was not asked of other foods to test this hypothesis further.

Eating chicken outside the home was identified as a risk factor in the univariate analysis (Table 7.6). However, although some of the locations where it was consumed (e.g. restaurant) had an OR>1.0 they were not statistically significant. This risk factor was not found to be significant in any of the multivariate models.

**Foods eaten - poultry other than chicken**

As mentioned above eating poultry other than chicken prepared at home was "protective" as was handling it in the kitchen (Table 7.6). The same argument as previously given applies here.

**Foods from other animals**

The only significant risk factor from the univariate analysis was eating lamb (P=0.049). It is well known that sheep excrete *Campylobacter* (Ogden, Dallas et al. 2009) and so it is plausible for lamb to be a risk factor. The numbers were relatively small (33 cases and 6 controls) and none of the ways that the lamb was prepared gave statistically significant results. Lamb remained a risk factor in the unweighted multivariate models (Table 7.7 and Table A7.3) but was not found to be significant in the multivariate weighted analysis (Table 7.8 and Table A7.4)

The above analysis has generated some potential risk factors associated with foreign travel acquired campylobacteriosis. However, none of the risk factors remain significant across the univariate analysis and the range of multivariate models developed. This is likely due to the relatively small number of questionnaires being analysed and the range of destinations visited and activities which individuals carry out which likely lead to a wide spectrum of exposures and hence mechanisms of infection.

The percentage of cases and controls correctly assigned by the multivariate models varied between 74.2% and 81.2% (Table 7.5). As expected the models with P<0.157 had increased classification accuracy because more variables were contained within the final model.

**Foreign Travel by Region of Destination**

Table 7.9 presents the odds ratios and P-values by region of destination. Visiting Asia is a risk factor (OR 10.20, P=0.026). Asia had also been found to be a risk factor in previous studies on returning travellers from The Netherlands (Mughini-Gras, Smid et al. 2014), Sweden (Indian subcontinent and Turkey (Ekdahl, Andersson 2004) and the Aberdeen City and Shire study (Smith-Palmer, Cowden 2010). The risk of campylobacteriosis was found to be lower when travelling to North America (OR = 0.27, P=0.044). The Aberdeen city and Shire study also found the risk decreased when travelling to North America (Canada and USA) but this was not statistically significant (Smith-Palmer, Cowden 2010).

**Table 7.9 Univariate analysis of Foreign Travel cases by Region of Destination**

| Region of Destination | Cases exposed[a] N (%) | Controls exposed[b] N (%) | Adjusted OR[c] (95% CI) | P value |
|---|---|---|---|---|
| Western Europe | 23 (15.8) | 15 (28.8) | 0.47 (0.22 - 1.04) | 0.061 |
| Eastern Europe | 11 (7.5) | 3 (5.8) | 1.38 (0.35 - 5.38) | 0.647 |
| Northern Europe | 4 (2.7) | 3 (5.8) | 0.51 (0.10 - 2.49) | 0.406 |
| Southern Europe | 76 (52.1) | 21 (40.4) | 1.65 (0.84 - 3.24) | 0.146 |
| Africa | 12 (8.2) | 4 (7.7) | 0.96 (0.28 - 3.23) | 0.942 |
| Asia | 25 (17.1) | 1 (1.9) | 10.20 (1.33 - 76.92) | 0.026 |
| Australasia | 1 (0.7) | 1 (1.9) | 0.22 (0.01 - 3.89) | 0.301 |
| North America | 5 (3.4) | 6 (11.5) | 0.27 (0.08 - 0.97) | 0.044 |
| South America | 1 (0.7) | 0 (0.0) | nd | |

[a] There were 146 cases that travelled abroad. In the table the sum of cases totals 158. That is because 12 individuals travelled to two of the destinations listed above.
[b] There were 52 controls (i.e. 52 campylobacteriosis cases that did not travel abroad). The sum of controls totals 54 because 2 individuals travelled to two of the destinations listed above.
[c] Adjusted by age and sex

## 7.3 Estimate of the contribution of foreign travel to the difference in reported campylobacteriosis between the SIMD5 and SIMD1 using case and control data

The excess in the incidence of human campylobacteriosis in the SIMD5 compared with the SIMD1 populations can be seen in Figure 4.2. This corresponds to an average number of reported cases in this period (2012-2017) of 1218 from SIMD5 and 957 from SIMD1 per year. The difference being 261 cases.

From the case control study it is known that foreign travel is associated with 27.0% (123/455) of cases in SIMD5 and 16.3% (22/135) in SIMD1. By calculating the odds ratios and these percentages it is possible to determine the population attributable fractions for SIMD1 and SIMD5.

For SIMD5 the odds ratio was calculated to be 2.98 (95% CI 2.05 to 4.32) and the proportion of cases exposed is 0.27. Hence, the PAF is 18.0%. For SIMD1 the odds ratio was 5.22 (95% CI 1.91 to 14.2) and the proportion of cases exposed is 0.163. Hence, the PAF is 13.2%. Hence, the number of foreign travel associated

cases in SIMD5 is 0.180*1218 = 218.8. Also the number of foreign travel associated cases from SIMD1 is 0.132*957 = 126.1. The excess of approximately 92.7 cases (=218.8-126.1) in SIMD5 is therefore attributable to foreign travel. The proportional excess in SIMD5 is thus 92.7/261 or 36% (95% CI 15.0 to 59.0%). The CIs were calculated by propagating the standard error in the OR's through the equations.

This finding indicates that a large part of the difference in campylobacteriosis incidence between SIMD5 and SIMD1 is due to foreign travel. Figure 4.2 also shows that the incidence of campylobacteriosis is also higher in SIMD2, SIMD3 and SIMD4 compared with the most deprived SIMD1 quintile. It may be that these differences are similarly substantially due to foreign travel. However, to confirm this would require foreign travel information on cases and controls from these areas.

## 7.4 Conclusions

### 7.4.1 Domestic Case Control Study

The case-control study found that cases from least deprived SIMD5 areas were more likely to report campylobacteriosis than those from most deprived SIMD1 areas but this was not statistically significant. The only socioeconomic factor that was associated with an increased risk of campylobacteriosis was having a car <5 years old. This was statistically significant in all of the analysis conducted. None of the factors associated with deprivation (e.g. being on benefits or unemployed) were significant in the analysis.

Eating chicken liver pâté prepared at home (PAF =5%), eating chicken lightly cooked (PAF =2 to 3%), eating chicken outside the home elsewhere (not restaurant, take-away or fast food) (PAF=8 to 9%) and eating poultry (other than chicken) at a restaurant (PAF=5%) were all significant risk factors in all of the analysis. The following non-food risk factors were also consistently significant in all of the analysis: being on PPIs (PAF 10 to 16%) and having white ethnicity (white) (PAF 71 to 77%). However, chicken is a complex risk factor and can be "protective" depending on the setting and where it is prepared. For example, raw chicken handled in the kitchen was "protective" which was counter-intuitive.

Using an indoor swimming pool/toddler pool was protective in all of the analysis. Contact with various animal groups, predominantly pets was mostly protective but results were not totally consistent as was consumption of a number of other foods (e.g. pork and beef).

### 7.4.2 Foreign Travel Case Control Study

Foreign travel is a risk factor for human campylobacteriosis and there were proportionally 3 times as many cases associated with travel abroad than not. Increased risk of campylobacteriosis occurred when travelling to Asia (including Turkey) and lower risk for travelling to North America.

The study also found that there was statistically no difference of campylobacteriosis risk when comparing cases and controls between most and least deprived areas. However, there were very few completed questionnaires from cases (22) and controls (5) in the most deprived (SIMD1) areas.

The univariate analysis found two risk factors for foreign travel associated campylobacteriosis (eating chicken outside the home and eating lamb) but only the unweighted multivariate models found eating lamb as a risk factor. A number of factors were found to reduce the risk (e.g. household income >£47k per year, H2 blockers, dog contact and eating chicken prepared at home) but none of these were consistently significant across the models. The small numbers of questionnaires completed (146 cases and 52 controls) will have reduced the statistical power of the analysis.

### 7.4.3 Impact of foreign travel on differential reported incidence rates for SIMD1 and SIMD5.

Between 2012 and 2017 on average each year there were 957 cases reported in SIMD1 areas and 1218 in SIMD5. The difference in reported cases averaged 261 of which approximately 36% could be explained by greater frequency of foreign travel of the SIMD5 population.

# 8. Case-case Analysis

## 8.1 Introduction

Case-control studies have received some criticism because of potential biases between selection of cases and controls (McCarthy, Giesecke 1999). For example, in diseases such as campylobacteriosis, where only a fraction of cases are reported, those that are reported may be non-random because of the operation of the surveillance system compared with a randomly selected group of controls.

Case-case analysis has been developed which reduce such biases but may lead to the development of new biases. For example if a case-case analysis was performed between two infectious diseases and if the main vehicle of infection was the same for both diseases then the analysis may not identify this as a risk factor as it is the same for both diseases.

A number of case-case studies have been carried out comparing human campylobacteriosis with another disease or subdividing campylobacteriosis by species (*C. jejuni* and *C. coli*) or subtype (e.g. by MLST). In the British Columbia region of Canada a case-case study was conducted comparing campylobacteriosis with other reported enteric diseases. These diseases were therefore recorded using the same surveillance system. It found that campylobacteriosis was more common than enteric diseases for cases served by private wells, living in rural settings, aged greater than 15 years and higher socioeconomic status (Galanis, Mak et al. 2014).

A Scottish case-case study comparing risk factors for *C. coli* and *C. jejuni* infection found that there was a higher risk of contracting *C. coli* infection in the summer and in people >19 years of age whilst the risk was reduced when living in an urban area (Roux, Sproston et al. 2013).

Case-case analysis can also be used to identify differences in risk factors between different populations who suffer from the same disease. In Arizona, USA, such an analysis has been done comparing cases of campylobacteriosis between Hispanic and Non-Hispanic populations (Pogreba-Brown, Barrett 2018). Differences in age, rurality, seasonality and disease presentation were found between the ethnic groups. It was also found that Hispanics had a higher likelihood of consuming higher risk foods (e.g. queso fresco, cilantro and animal products) whilst Non-Hispanic groups had a greater risk of environmental exposure.

Here the case-case methodology will be used to compare cases from most deprived (SIMD1) and least deprived (SIMD5) populations. This will utilise the case data from the case-control questionnaire. The case-case analysis will be conducted in two ways. The first will utilise all of the factors employed in the domestic case-control study (note that foreign travel cases are excluded from this analysis). This will include personal characteristic data which would be expected to be associated with the Scottish Index of Multiple Deprivation. For example, the socioeconomic variables, being on benefits and unemployed would be expected to be associated with data zones which are in SIMD1, whilst high

income and living in a house >3 bedrooms would be expected to be associated with SIMD1. It is highly likely that these factors will dominate the analysis, but they will not be that informative about explaining the differences in how and why individuals contract campylobacteriosis between these two populations. They will however provide some evidence on the distribution of "rich" and "poor" people in SIMD1 and SIMD5 data zones and go some way in identifying whether the ecological fallacy is likely to be important or not. The ecological fallacy explains why ecological studies performed at a population level (e.g. data zone populations) may not be representative of each individual in a population. For example there may be some affluent people living in deprived areas and vice-versa (Haneuse, Wakefield 2007).

The second approach will remove these variables from the analysis to see whether other factors (food, water, animal contact etc.) are differentially associated between cases from SIMD1 and SIMD5. This approach is likely to be more valuable as these factors are more likely to be modifiable in the short term (e.g. it should be easier to inform people about what they can do to reduce a foodborne risk whilst it is much harder to reduce the number of people on benefits).

## 8.2 Perform case-case analysis using logistic regression

### 8.2.1 Data

The data extracted from the case questionnaires were utilised in this study. This included general details about the individual (age etc.), details of household income, historical health conditions, domestic travel and exposure to animals, food and water. The data were split into least deprived (SIMD5) and most deprived groups (SIMD1). In total there were 332 SIMD5 and 113 SIMD1 cases respectively (See Figure 6.1).

### 8.2.2 Methods

#### 8.2.2.1 Descriptive analysis

For each factor the number of SIMD5 cases and SIMD1 cases exposed were determined as well as the number of SIMD5 cases and SIMD1 cases where data were incomplete.

#### 8.2.2.2 Univariate and multivariate logistic regression

Univariate regression adjustments: For the case-case analysis the following confounding variables (season, age, sex and rurality) which had been previously identified as risk factors for human campylobacteriosis (Kuhn, Nielsen et al. 2018) were assessed by univariate logistic regression (SPSS 25) to determine whether they were significant ($P<0.05$) risk factors. For those variables that were statistically significant adjustments were made in both the univariate and multivariate analysis.

Univariate and multivariate regression case-case analysis: The univariate analysis was performed for all explanatory variables (risk factors) utilising logistic regression which generated adjusted ORs and 95% confidence intervals. Variables with a p-value of <0.25 were candidates for the multivariate analyses. Two approaches were taken on the selection of variables. The first utilised all variables with a p-value <0.25. The second removed the main socioeconomic variables (being on benefits, household income >£47k per year, retired, school aged, student, unemployed, professional job, car <5 years old, house >= 3 bedrooms, sit/sat on a committee or council, profession (other)).

Multivariate logistic regression was performed by backwards stepwise elimination with non-significant variables removed one step at a time. Missing data were inferred by multiple imputation with 100 iterations and a pooled model was generated. Backwards step elimination was repeated until only variables were left with P<0.157 and P<0.05. These two models were kept for further analysis.

Multivariate models were performed where the data were both unweighted and weighted to correct for sample bias (Chapter 6 sections 6.8 and 6.9). Each model was tested to determine how many of the cases and controls were correctly assigned.

## 8.2.3 Results and discussion

### 8.2.3.1 Case-Case logistic regression analysis

Adjustments: Table 8.1 presents the results of the univariate analysis for those variables selected as possible adjustments for confounding. There was proportionally more cases in the 15-24 and 25-64 age groups in the most deprived population (i.e. SIMD1) compared with the reference 65+ age group. In contrast there were more peri-urban cases in the least deprived population (i.e. SIMD5) compared with urban areas. There were no significant differences obtained for the season and sex variables. As such it was decided to use age and rurality as adjustment factors in the subsequent univariate and multivariate analysis.

**Table 8.1 Univariate analysis of potential adjusting variables for case-case study**

| **Domestic -Case Control Study** | | | | | | |
|---|---|---|---|---|---|---|
| Characteristic | Cases SIMD5 (N=332) n (%) | Cass SIMD5 Unk | Cases SIMD1 (N=113) n (%) | Cases SIMD1 Unk. | OR (95% CI) | P-value |
| **Season** | | | | | | |
| Summer | 160 (48.2) | 0 | 54 (47.8) | 0 | 1.016 (0.663 – 1.558) | 0.941 |
| Rest of Year (Ref.) | 172 (51.8) | 0 | 59 (52.2) | 0 | 1 | |
| **Age** | | | | | | |
| 5 - 14 | 10 (3.0) | 17 | 0 (0) | 1 | ∞ | |
| 15 - 24 | 16 (4.8) | 17 | 10 (8.8) | 1 | 0.311 (0.124 – 0.779) | 0.013 |
| 25 - 64 | 181 (54.5) | 17 | 81 (71.7) | 1 | 0.434 (0.254 – 0.743) | 0.002 |
| 65+ (Ref.) | 108 (32.5) | 17 | 21 (18.6) | 1 | 1 | |
| **Sex** | | | | | | |
| Male | 179 (53.9) | 0 | 53 (46.9) | 0 | 1.324 (0.863 – 2.032) | 0.198 |
| Female (Ref.) | 153 (46.1) | 0 | 60 (53.1) | 0 | 1 | |
| **Rurality** | | | | | | |
| Rural | 23 (6.9) | 0 | 8 (7.1) | 0 | 1.174 (0.506 – 2.723) | 0.709 |
| Peri-urban | 91 (27.4) | 0 | 16 (14.2) | 0 | 2.322 (1.293 – 4.171) | 0.005 |
| Urban (Ref.) | 218 (65.7) | 0 | 89 (78.8) | 0 | 1 | |

Factors which were more common in the least deprived part of the population (SIMD5) are coloured red. Those more common in the most deprived part of the population (SIMD1) are coloured blue.

Univariate Analysis: Table 8.2 presents the results of the univariate analysis with adjustments. There were 10 factors that were more common in the least deprived part of the population (SIMD5 coloured red in Table 8.2) whilst 11 were found more common in the most deprived part of the population (SIMD1, coloured blue in Table 8.2). When looking at the personal characteristics those variables expected to be associated with areas of most deprivation (e.g. being on benefits and unemployed) were significant. Whereas those associated with wealth (e.g. household income >£47k per year, professional job, car <5 years old and house > 3 bedrooms) were more common in the least deprived (SIMD5)

areas. However not all unemployed people lived in the most deprived areas (2.7% of the least deprived cases were unemployed) and not all people with incomes >£47k lived in the least deprived areas (13% of cases in the most deprived areas had >£47k income).

**Table 8.2 Univariate analysis of Risk Factors for case-case study**

| General Details - Personal characteristics | Cases SIMD5, n (%) | Cases SIMD5 Unknowns | Cases SIMD1, n (%) | Cases SIMD1 Unknowns | Adjusted OR (95% CI) | P-value |
|---|---|---|---|---|---|---|
| **N** | 332 | | 113 | | | |
| **Season** | | | | | | |
| Summer | 160 (48.2) | 0 | 54 (47.8) | 0 | 1.12 (0.71-1.74) | 0.633 |
| Rest of year (Reference) | 172 (51.8) | 0 | 59 (52.2) | 0 | | |
| **Age (years)** | | | | | | |
| 5-14 | 10 (3.0) | 17 | 0 (0) | 1 | ∞ | |
| 15-24 | 16 (4.8) | 17 | 10 (8.8) | 1 | 0.3 (0.12-0.76) | 0.011 |
| 25-64 | 181 (54.5) | 17 | 81 (71.7) | 1 | 0.43 (0.25-0.74) | 0.002 |
| 65+ (Reference) | 108 (32.5) | 17 | 21 (18.6) | 1 | | |
| **Gender** | | | | | | |
| Male | 179 (53.9) | 0 | 53 (46.9) | 0 | 1.16 (0.74-1.82) | 0.524 |
| Female (Reference) | 153 (46.1) | 0 | 60 (53.1) | 0 | | |
| **Ethnicity** | | | | | | |
| White | 329 (99.1) | 0 | 111 (98.2) | 0 | 1.35 (0.21-8.53) | 0.747 |
| Other (Reference) | 3 (0.9) | 0 | 2 (1.8) | 0 | | |

| (continued) | Cases SIMD5, n (%) | Cases SIMD5 Unknowns | Cases SIMD1, n (%) | Cases SIMD1 Unknowns | Adjusted OR (95% CI) | P-value |
|---|---|---|---|---|---|---|
| **SIMD** | | | | | | |
| SIMD1 | 0 (0) | 0 | 113 (100) | 0 | nd | nd |
| SIMD5 (Reference) | 332 (100) | 0 | 0 (0) | 0 | | |
| **Rurality, benefits and income** | | | | | | |
| Rural | 23 (6.9) | 0 | 8 (7.1) | 0 | 1.15 (0.49-2.72) | 0.747 |
| peri-Urban | 91 (27.4) | 0 | 16 (14.2) | 0 | 2.38 (1.29-4.39) | 0.006 |
| Urban (Reference) | 218 (65.7) | 0 | 89 (78.8) | 0 | | |
| Benefits | 20 (6) | 7 | 45 (39.8) | 3 | 0.09 (0.04-0.16) | <0.001 |
| Household income >£47k/year | 143 (43.1) | 19 | 15 (13.3) | 5 | 6.75 (3.64-12.53) | <0.001 |
| **Occupation** | | | | | | |
| Retired | 134 (40.4) | 0 | 22 (19.5) | 0 | 2.41 (1.19-4.91) | 0.015 |
| School aged | 13 (3.9) | 0 | 3 (2.7) | 0 | 1.6 (1.6-0.28) | 0.595 |
| Student | 11 (3.3) | 0 | 6 (5.3) | 0 | 0.67 (0.2-2.23) | 0.509 |
| Unemployed | 9 (2.7) | 0 | 13 (11.5) | 0 | 0.2 (0.07-0.54) | 0.001 |
| Professional job | 134 (40.4) | 0 | 27 (23.9) | 0 | 4.17 (2.41-7.23) | <0.001 |
| Car <5 years old | 189 (56.9) | 0 | 36 (31.9) | 0 | 3.08 (1.92-4.94) | <0.001 |
| House >= 3 bedrooms | 244 (73.5) | 0 | 45 (39.8) | 0 | 4.34 (2.7-6.96) | <0.001 |

| (continued) | Cases SIMD5, n (%) | Cases SIMD5 Unknowns | Cases SIMD1, n (%) | Cases SIMD1 Unknowns | Adjusted OR (95% CI) | P-value |
|---|---|---|---|---|---|---|
| Sit/Sat on a committee or council | 55 (16.6) | 0 | 5 (4.4) | 0 | 3.25 (1.23-8.57) | 0.017 |
| Profession - None of the above | 12 (3.6) | 0 | 29 (25.7) | 0 | 0.1 (0.05-0.21) | <0.001 |
| **Historical health conditions and treatment** | | | | | | |
| Long term bowel condition | 47 (14.2) | 0 | 28 (24.8) | 0 | 0.39 (0.22-0.7) | 0.001 |
| Other medical condition | 132 (39.8) | 0 | 45 (39.8) | 0 | 0.81 (0.51-1.29) | 0.382 |
| PPIs | 71 (21.4) | 0 | 31 (27.4) | 0 | 0.64 (0.39-1.08) | 0.093 |
| H2-blockers | 6 (1.8) | 0 | 6 (5.3) | 0 | 0.32 (0.1-1.04) | 0.058 |
| Antacids | 33 (9.9) | 0 | 20 (17.7) | 0 | 0.49 (0.26-0.92) | 0.027 |
| Antiflatuents | 7 (2.1) | 0 | 5 (4.4) | 0 | 0.58 (0.17-1.95) | 0.379 |
| PPIs, H2-blockers, Antacids, Antiflatuents | 102 (30.7) | 0 | 50 (44.2) | 0 | 0.52 (0.33-0.83) | 0.005 |
| Antibiotics | 20 (6.0) | 0 | 9 (8.0) | 0 | 0.92 (0.4-2.15) | 0.854 |
| Medicine (other) | 145 (43.7) | 0 | 55 (48.7) | 0 | 0.71 (0.45-1.11) | 0.138 |
| **Travel** | | | | | | |
| Travel within Scotland | 64 (19.3) | 5 | 15 (13.3) | 2 | 1.73 (0.91-3.28) | 0.092 |
| Travel outside Scotland (not abroad) | 34 (10.2) | 25 | 6 (5.3) | 16 | 2.04 (0.81-5.13) | 0.131 |
| Foreign travel | 0 (0) | 28 | 0 (0) | 17 | nd | nd |

| (continued) | Cases SIMD5, n (%) | Cases SIMD5 Unknowns | Cases SIMD1, n (%) | Cases SIMD1 Unknowns | Adjusted OR (95% CI) | P-value |
|---|---|---|---|---|---|---|
| **Contact with animals** | | | | | | |
| Contact with animals -overall | 172 (51.8) | 1 | 53 (46.9) | 2 | 1.32 (0.83-2.08) | 0.239 |
| Dogs | 130 (39.2) | 0 | 39 (34.5) | 0 | 1.28 (0.8-2.04) | 0.298 |
| Cats | 55 (16.6) | 0 | 17 (15.0) | 0 | 1.47 (0.79-2.74) | 0.224 |
| Birds/Poultry | 8 (2.4) | 0 | 8 (7.1) | 0 | 0.3 (0.1-0.91) | 0.034 |
| Farm animals(cattle, sheep, goats, horses, donkeys, pigs) | 7 (2.1) | 0 | 3 (2.7) | 0 | 0.79 (0.18-3.49) | 0.759 |
| Other animals | 11 (3.3) | 0 | 6 (5.3) | 0 | 0.71 (0.24-2.09) | 0.532 |
| Contact with ill animal | 7 (2.1) | 0 | 3 (2.7) | 0 | 0.79 (0.19-3.25) | 0.747 |
| Touch animal faeces | 27 (8.1) | 27 | 9 (8.0) | 15 | 0.91 (0.39-2.09) | 0.817 |
| **Exposure to water** | | | | | | |
| Water activity | 35 (10.5) | 3 | 4 (3.5) | 2 | 3.07 (1.03-9.14) | 0.044 |
| Indoor swimming pool / toddler pool | 22 (6.6) | 3 | 2 (1.8) | 2 | 3.6 (0.8-16.14) | 0.094 |
| Outdoor swimming pool / paddling pool / theme park water ride / splash park | 2 (0.6) | 3 | 0 (0) | 2 | ∞ | |
| Loch/lake/pond/stream/river/burn(e.g. swimming, canoeing, diving, fishing) | 5 (1.5) | 3 | 1 (0.9) | 2 | 1.92 (0.2-18.36) | 0.572 |

| (continued) | **Cases SIMD5, n (%)** | **Cases SIMD5 Unknowns** | **Cases SIMD1, n (%)** | **Cases SIMD1 Unknowns** | **Adjusted OR (95% CI)** | **P-value** |
|---|---|---|---|---|---|---|
| Sea(e.g. diving, sailing, surfing, jet ski, fishing) | 3 (0.9) | 3 | 0 (0) | 2 | ∞ | 0.999 |
| Other water activity | 7 (2.1) | 3 | 1 (0.9) | 2 | 2.62 (0.31-22.12) | 0.377 |
| Water source (public mains) | 321 (96.7) | 5 | 98 (86.7) | 1 | 7.3 (2.66-20.06) | <0.001 |
| Water source (private-spring) | 2 (0.6) | 5 | 2 (1.8) | 1 | 0.25 (0.03-1.95) | 0.186 |
| Water source (private-well) | 2 (0.6) | 5 | 3 (2.7) | 1 | 0.2 (0.03-1.39) | 0.104 |
| Water source (River/stream/lake/loch/pond/melted snow (not boiled) | 3 (0.9) | 5 | 0 (0) | 1 | ∞ | 0.999 |
| **Exposure to food** | | | | | | |
| **Food Eaten - Chicken** | | | | | | |
| Eat chicken prepared at home | 200 (60.2) | 20 | 62 (54.9) | 5 | 1.58 (0.98-2.54) | 0.062 |
| Eat chicken outside the home | 109 (33) | 0 | 37 (33.0) | 0 | 1.09 (0.44-2.71) | 0.850 |
| Chicken outside the home - Restaurant | 48 (14.5) | 0 | 13 (11.5) | 0 | 1.29 (0.65-2.55) | 0.465 |
| Chicken outside the home - Take away or Fast food | 45 (13.6) | 0 | 20 (17.7) | 0 | 0.89 (0.48-1.64) | 0.707 |
| Chicken outside the home - Elsewhere | 60 (18.1) | 0 | 24 (21.2) | 0 | 0.86 (0.49-1.51) | 0.594 |

| (continued) | **Cases SIMD5, n (%)** | **Cases SIMD5 Unknowns** | **Cases SIMD1, n (%)** | **Cases SIMD1 Unknowns** | **Adjusted OR (95% CI)** | **P-value** |
|---|---|---|---|---|---|---|
| Eat chicken liver pâté prepared from raw at home | 23 (6.9) | 0 | 7 (6.2) | 0 | 1.09 (0.44-2.71) | 0.850 |
| Eat chicken liver pâté prepared outside the home | 5 (1.5) | 0 | 1 (0.9) | 0 | 1.5 (0.16-14.15) | 0.722 |
| Frozen chicken purchased which was then prepared at home | 65 (19.6) | 36 | 22 (19.5) | 11 | 1.05 (0.59-1.86) | 0.867 |
| Fresh raw chicken purchased which was then prepared at home | 180 (54.2) | 23 | 62 (54.9) | 8 | 1.12 (0.7-1.8) | 0.641 |
| Raw chicken washed before preparation | 37 (11.1) | 38 | 24 (21.2) | 16 | 0.5 (0.28-0.91) | 0.024 |
| Raw chicken cut up in the kitchen | 122 (36.7) | 38 | 38 (33.6) | 16 | 1.39 (0.84-2.29) | 0.197 |
| Raw chicken handled in the kitchen | 109 (32.8) | 38 | 35 (31.0) | 17 | 1.22 (0.74-2) | 0.439 |
| Raw chicken at home - oven-cooked, roasted or grilled | 102 (30.7) | 32 | 34 (30.1) | 16 | 1.07 (0.65-1.76) | 0.801 |
| Chicken outside the home - oven-cooked, roasted or grilled | 72 (21.7) | 0 | 20 (17.7) | 0 | 1.37 (0.76-2.45) | 0.295 |
| Raw chicken at home - BBQ | 18 (5.4) | 0 | 1 (0.9) | 0 | 4.79 (0.62-37.21) | 0.134 |
| Chicken outside the home - BBQ | 13 (3.9) | 0 | 4 (3.5) | 0 | 1.02 (0.3-3.4) | 0.978 |
| Raw chicken at home - stir fried | 50 (15.1) | 32 | 19 (16.8) | 16 | 0.96 (0.52-1.78) | 0.907 |
| Chicken outside the home - stir fried | 13 (3.9) | 0 | 6 (5.3) | 0 | 0.95 (0.34-2.67) | 0.928 |

| (continued) | **Cases SIMD5, n (%)** | **Cases SIMD5 Unknowns** | **Cases SIMD1, n (%)** | **Cases SIMD1 Unknowns** | **Adjusted OR (95% CI)** | **P-value** |
|---|---|---|---|---|---|---|
| Raw chicken at home – microwaved | 2 (0.6) | 32 | 0 (0) | 16 | ∞ | |
| Chicken outside the home - microwaved | 4 (1.2) | 0 | 3 (2.7) | 0 | 0.43 (0.08-2.16) | 0.303 |
| Raw chicken at home - stewed, slow cooked or steamed | 26 (7.8) | 32 | 5 (4.4) | 16 | 1.73 (0.63-4.76) | 0.285 |
| Chicken outside the home - stewed, slow cooked or steamed | 10 (3.0) | 0 | 4 (3.5) | 0 | 0.78 (0.22-2.75) | 0.783 |
| Raw chicken at home - deep fried | 2 (0.6) | 32 | 1 (0.9) | 16 | 0.75 (0.07-8.72) | 0.821 |
| Chicken outside the home - deep fried | 13 (3.9) | 0 | 9 (8.0) | 0 | 0.67 (0.26-1.7) | 0.398 |
| Chicken lightly cooked (i.e. pinkish in the middle) | 10 (3.0) | 70 | 1 (0.9) | 26 | 3.01 (0.37-24.52) | 0.304 |
| **Foods Eaten – Poultry other than chicken** | | | | | | |
| Eat poultry other than chicken prepared at home | 20 (6.0) | 25 | 8 (7.1) | 4 | 0.89 (0.37-2.16) | 0.800 |
| Eat poultry other than chicken prepared outside the home | 25 (7.5) | 0 | 10 (8.8) | 0 | 0.76 (0.34-1.7) | 0.501 |
| Poultry (other than chicken) outside the home - Restaurant | 20 (6.0) | 0 | 5 (4.4) | 0 | 1.09 (0.38-3.13) | 0.875 |
| Poultry (other than chicken) outside the home - Take away or Fast food | 2 (0.6) | 0 | 3 (2.7) | 0 | 0.37 (0.06-2.29) | 0.286 |

| (continued) | Cases SIMD5, n (%) | Cases SIMD5 Unknowns | Cases SIMD1, n (%) | Cases SIMD1 Unknowns | Adjusted OR (95% CI) | P-value |
|---|---|---|---|---|---|---|
| Poultry (other than chicken) outside the home - Elsewhere | 14 (4.2) | 0 | 6 (5.3) | 0 | 0.68 (0.24-1.9) | 0.458 |
| Eat poultry liver pâté (other than chicken) which was prepared from raw at home | 2 (0.6) | 28 | 1 (0.9) | 5 | 0.64 (0.05-7.53) | 0.724 |
| Eat poultry liver pâté (other than chicken) prepared outside home | 6 (1.8) | 0 | 1 (0.9) | 0 | 2.42 (0.28-20.79) | 0.422 |
| Frozen poultry (other than chicken) purchased which was then prepared at home | 4 (1.2) | 30 | 6 (5.3) | 8 | 0.25 (0.07-0.95) | 0.042 |
| Fresh raw poultry (other than chicken) purchased which was then prepared at home | 14 (4.2) | 41 | 6 (5.3) | 10 | 0.77 (0.28-2.17) | 0.628 |
| Raw poultry (other than chicken) washed before preparation | 5 (1.5) | 53 | 1 (0.9) | 13 | 1.77 (0.19-16.35) | 0.615 |
| Raw poultry (other than chicken) cut up in the kitchen | 6 (1.8) | 53 | 4 (3.5) | 13 | 0.5 (0.12-1.97) | 0.320 |
| Raw poultry (other than chicken) handled in the kitchen | 7 (2.1) | 53 | 2 (1.8) | 13 | 1.13 (0.22-5.85) | 0.886 |
| Raw poultry (other than chicken) at home - oven-cooked, roasted or grilled | 11 (3.3) | 53 | 2 (1.8) | 13 | 1.75 (0.36-8.42) | 0.484 |

| (continued) | **Cases SIMD5, n (%)** | **Cases SIMD5 Unknowns** | **Cases SIMD1, n (%)** | **Cases SIMD1 Unknowns** | **Adjusted OR (95% CI)** | **P-value** |
|---|---|---|---|---|---|---|
| Poultry (other than chicken) outside the home - oven-cooked, roasted or grilled | 17 (5.1) | 0 | 8 (7.1) | 0 | 0.68 (0.28-1.70) | 0.413 |
| Raw poultry (other than chicken) at home - BBQ | 2 (0.6) | 53 | 0 (0) | 13 | ∞ | |
| Poultry (other than chicken) outside the home - BBQ | 5 (1.5) | 0 | 0 (0) | 0 | ∞ | |
| Raw poultry (other than chicken) at home - stir fried | 1 (0.3) | 53 | 2 (1.8) | 13 | 0.28 (0.02-3.23) | 0.310 |
| Poultry (other than chicken) outside the home - stir fried | 2 (0.6) | 0 | 2 (1.8) | 0 | 0.53 (0.07-3.82) | 0.526 |
| Raw poultry (other than chicken) at home - microwaved | 0 (0) | 53 | 0 (0) | 13 | ∞ | |
| Poultry (other than chicken) outside the home - microwaved | 2 (0.6) | 0 | 0 (0) | 0 | ∞ | |
| Raw poultry (other than chicken) at home - stewed, slow cooked or steamed | 0 (0) | 53 | 2 (1.8) | 13 | 0 | |
| Poultry (other than chicken) outside the home - stewed, slow cooked or steamed | 4 (1.2) | 0 | 0 (0) | 0 | ∞ | |

| (continued) | Cases SIMD5, n (%) | Cases SIMD5 Unknowns | Cases SIMD1, n (%) | Cases SIMD1 Unknowns | Adjusted OR (95% CI) | P-value |
|---|---|---|---|---|---|---|
| Raw poultry (other than chicken) at home - deep fried | 0 (0) | 0 | 0 (0) | 0 | nd | nd |
| Poultry (other than chicken) outside the home - deep fried | 3 (0.9) | 0 | 2 (1.8) | 0 | 0.61 (0.09-3.95) | 0.601 |
| Poultry (other than chicken) lightly cooked (i.e. pinkish in the middle) | 2 (0.6) | 53 | 0 (0) | 13 | ∞ | |
| **Foods Eaten - Other** | | | | | | |
| Eat either beef, pork, lamb, deer or rabbit | 245 (73.8) | 20 | 69 (61.1) | 9 | 1.81 (1.09-3.02) | 0.022 |
| Eat beef | 209 (63.0) | 20 | 58 (51.3) | 9 | 1.54 (0.96-2.48) | 0.072 |
| Eat beef undercooked (i.e. pinkish in the middle) | 44 (13.3) | 0 | 8 (7.1) | 0 | 2.19 (0.98-4.89) | 0.056 |
| Eat beef oven-cooked, roasted or grilled | 122 (36.7) | 20 | 28 (24.8) | 9 | 1.85 (1.11-3.08) | 0.019 |
| Eat beef BBQ | 18 (5.4) | 20 | 6 (5.3) | 9 | 1.07 (0.4-2.88) | 0.898 |
| Eat beef stir fried | 22 (6.6) | 20 | 8 (7.1) | 9 | 0.9 (0.38-2.17) | 0.820 |
| Eat beef microwaved | 5 (1.5) | 20 | 1 (0.9) | 9 | ∞ | |
| Eat beef deep fried | 2 (0.6) | 20 | 1 (0.9) | 9 | 0.4 (0.03-5.47) | 0.493 |
| Eat beef stewed, slow cooked or steamed | 95 (28.6) | 20 | 27 (23.9) | 9 | 1.12 (0.66-1.89) | 0.667 |

| (continued) | **Cases SIMD5, n (%)** | **Cases SIMD5 Unknowns** | **Cases SIMD1, n (%)** | **Cases SIMD1 Unknowns** | **Adjusted OR (95% CI)** | **P-value** |
|---|---|---|---|---|---|---|
| Eat pork | 92 (27.7) | 20 | 33 (29.2) | 9 | 1 (0.61-1.65) | 0.997 |
| Eat pork undercooked (i.e. pinkish in the middle) | 5 (1.5) | 0 | 3 (2.7) | 0 | 0.66 (0.14-3.09) | 0.601 |
| Eat pork oven-cooked, roasted or grilled | 56 (16.9) | 20 | 24 (21.2) | 9 | 0.83 (0.47-1.46) | 0.518 |
| Eat pork BBQ | 9 (2.7) | 20 | 2 (1.8) | 9 | 1.91 (0.39-9.36) | 0.423 |
| Eat pork stir fried | 12 (3.6) | 20 | 2 (1.8) | 9 | 3.12 (0.67-14.49) | 0.146 |
| Eat pork microwaved | 3 (0.9) | 20 | 0 (0) | 9 | ∞ | |
| Eat pork deep fried | 3 (0.9) | 20 | 2 (1.8) | 9 | 0.35 (0.05-2.28) | 0.273 |
| Eat pork stewed, slow cooked or steamed | 18 (5.4) | 20 | 4 (3.5) | 9 | 1.42 (0.45-4.48) | 0.545 |
| Eat lamb | 48 (14.5) | 20 | 12 (10.6) | 9 | 1.24 (0.61-2.51) | 0.549 |
| Eat lamb undercooked (i.e. pinkish in the middle) | 8 (2.4) | 0 | 0 (0) | 0 | ∞ | |
| Eat lamb oven-cooked, roasted or grilled | 34 (10.2) | 20 | 10 (8.8) | 9 | 1.08 (0.5-2.34) | 0.851 |
| Eat lamb BBQ | 1 (0.3) | 20 | 0 (0) | 9 | ∞ | |
| Eat lamb stir fried | 2 (0.6) | 20 | 0 (0) | 9 | ∞ | |
| Eat lamb microwaved | 2 (0.6) | 20 | 0 (0) | 9 | ∞ | |

| (continued) | **Cases SIMD5, n (%)** | **Cases SIMD5 Unknowns** | **Cases SIMD1, n (%)** | **Cases SIMD1 Unknowns** | **Adjusted OR (95% CI)** | **P-value** |
|---|---|---|---|---|---|---|
| Eat lamb deep fried | 0 (0) | 20 | 0 (0) | 9 | nd | nd |
| Eat lamb stewed, slow cooked or steamed | 16 (4.8) | 20 | 2 (1.8) | 9 | 2.25 (0.49-10.48) | 0.300 |
| Eat deer or rabbit | 6 (1.8) | 20 | 2 (1.8) | 9 | 1.2 (0.23-6.4) | 0.829 |
| Eat deer or rabbit undercooked (i.e. pinkish in the middle) | 2 (0.6) | 0 | 1 (0.9) | 0 | 0.53 (0.03-8.61) | 0.655 |
| Eat deer or rabbit oven-cooked, roasted or grilled | 3 (0.9) | 20 | 0 (0) | 9 | ∞ | |
| Eat deer or rabbit BBQ | 1 (0.3) | 20 | 0 (0) | 9 | ∞ | |
| Eat deer or rabbit stir fried | 0 (0) | 20 | 0 (0) | 9 | ∞ | |
| Eat deer or rabbit microwaved | 0 (0) | 20 | 0 (0) | 9 | ∞ | |
| Eat deer or rabbit deep fried | 0 (0) | 20 | 0 (0) | 9 | ∞ | |
| Eat deer or rabbit stewed, slow cooked or steamed | 2 (0.6) | 20 | 2 (1.8) | 9 | 0.42 (0.05-3.39) | 0.417 |
| Eat raw or lightly cooked fish / shell fish / sea food (e.g. fish, crab, prawns, mussels, oysters, calamari, sushi etc) | 63 (19.0) | 21 | 20 (17.7) | 6 | 1.28 (0.71-2.32) | 0.406 |
| Eat any unpasteurised dairy products (incl. milk and cheese) | 34 (10.2) | 29 | 19 (16.8) | 9 | 0.58 (0.3-1.09) | 0.091 |

Factors which were more common in the least deprived part of the population (SIMD5) are coloured red. Those more common in the most deprived part of the population (SIMD1) are coloured blue.

**Multivariate Analysis – All variables:**

The unweighted multivariate analysis is provided in Table 8.3 (P<0.05) and Table A8.1 (P<0.157). Whilst the weighted multivariate analysis is presented in Table 8.4 (P<0.05) and Table A8.2 (P<0.157). Table 8.5 presents the percentage of cases correctly assigned (i.e. to SIMD1 or SIMD5) from the 100 imputations. It can be observed that between 82.8%-83.7% of the cases are correctly assigned from the four models.

### Table 8.3 Multivariate Analysis of Univariate Risk Factors (P<0.05) No Weights

| Factor | OR (95% CI) | P-value |
|---|---|---|
| Benefits | 0.13 (0.06 - 0.26) | <0.001 |
| Household income >£47k/year | 3.91 (1.98 - 7.71) | <0.001 |
| Retired | 2.24 (1.03 - 4.86) | 0.042 |
| House >= 3 bedrooms | 3.48 (1.98 - 6.12) | <0.001 |
| H2-blockers | 0.16 (0.04 - 0.71) | 0.016 |
| Water source (public mains) | 6.85 (1.83 - 25.69) | 0.004 |

Factors which were more common in the least deprived part of the population (SIMD5) are coloured red. Those more common in the most deprived part of the population (SIMD1) are coloured blue.

### Table 8.4 Multivariate Analysis of Univariate Risk Factors (P<0.05) with Weights

| Factor | OR (95% CI) | P-value |
|---|---|---|
| Benefits | 0.13 (0.06 - 0.27) | <0.001 |
| Household income >£47k/year | 4.22 (2.12 - 8.41) | <0.001 |
| Retired | 2.30 (1.08 - 4.92) | 0.032 |
| House >= 3 bedrooms | 4.39 (2.52 - 7.66) | <0.001 |
| H2-blockers | 0.16 (0.04 -0.64) | 0.010 |
| Antacids | 0.44 (0.21 -0.90) | 0.026 |
| Water source (public mains) | 8.03 (2.23 - 28.86) | 0.001 |

Factors which were more common in the least deprived part of the population (SIMD5) are coloured red. Those more common in the most deprived part of the population (SIMD1) are coloured blue.

**Table 8.5 Multivariate Model assignment**

| Dataset | Model | Percentage Correctly Assigned[a] |
|---|---|---|
| Case-Case – All variables | Multivariate model no weights (P<0.157) | 82.8 |
| | Multivariate model no weights (P<0.05) | 83.2 |
| | Multivariate model with weights (P<0.157) | 83.7 |
| | Multivariate model with weights (P<0.05) | 83.0 |
| Case-Case – Without the socioeconomic variables | Multivariate model no weights (P<0.157) | 76.7 |
| | Multivariate model no weights (P<0.05) | 76.7 |
| | Multivariate model with weights (P<0.157) | 78.3 |
| | Multivariate model with weights (P<0.05) | 78.3 |
| | Multivariate model with weights (P<0.157) | 78.3 |

[a] This is the percentage correctly assigned for the 100 imputations that were carried out in SPSS.

**General details/Personal characteristics:** Cases resident in the most deprived (SIMD1) areas are more likely to be in receipt of benefits in all of the multivariate models. Having a household income >£47k/year, being retired and having a house > 3 bedrooms was statistically significantly more common for cases in the least deprived (SIMD5) areas in all of the multivariate models. These results are unsurprising as they are in line with how SIMD1 and SIMD5 are defined.

**Historical Health Conditions and Treatment:** Taking H2 blockers is statistically significantly associated with cases living in the most deprived areas (SIMD1) for all of the models. Taking antacid was statistically significantly associated with living in the most deprived areas (SIMD1) but only for the weighted models and the univariate analysis.

**Travel History and Contact with animals:** None of the variables were statistically significant in the multivariate analysis.

**Exposure to water:** Having a public mains water source was associated with living in areas of least deprivation (SIMD5) in all of the models.

**Foods eaten:** Only one food exposure emerged as statistically significant in the multivariate analysis and that was only in the unweighted, P<0.157 model. Raw chicken was more likely to be washed before preparation for cases resident in the most deprived (SIMD1) areas (Table A8.2). It is unclear why this should be the case but could potentially be due to reduced exposure of consumer messaging not to wash chicken, or a cultural habit of washing chicken that is difficult to change.

**Vegetarian:** In the study there were 3 SIMD1 vegetarian cases (2.7%) compared to 12 in SIMD5 (3.6%) areas. This was not a statistically significant difference (OR = 1.37 (0.38 -4.96)).

**Multivariate Analysis – with socioeconomic variables removed:**

The multivariate analysis for the unweighted model is presented in Table 8.6 (P<0.05) and Table A8.3 (P<0.157). The weighted models are in Table 8.7 (P<0.05) and Table A8.4 (P<0.157). Table 8.5 presents the percentage of cases correctly assigned (i.e. to SIMD1 or SIMD5) from the 100 imputations. It can be observed that between 76.7%-78.3% of the cases are correctly assigned from the four weighted models. These percentages are lower than when the socioeconomic variables were included. This is to be expected because those variables should be associated with the level of deprivation and as such the models including them should have better assignment of cases.

### Table 8.6 Multivariate Analysis of Univariate Risk Factors (P<0.05) No Weights and sociodemographic variables removed

| Factor | OR (95% CI) | P-value |
|---|---|---|
| Long term bowel condition | 0.45 (0.25 - 0.83) | 0.01 |
| H2 Blockers | 0.26 (0.08 - 0.92) | 0.037 |
| Antacids | 0.47 (0.24 - 0.91) | 0.025 |
| Contact with Birds/Poultry | 0.31 (0.10 - 0.99) | 0.049 |
| Water source (public mains) | 5.09 (1.82 - 14.26) | 0.002 |
| Raw chicken washed before preparation | 0.39 (0.20 - 0.77) | 0.007 |

Factors which were more common in the least deprived part of the population (SIMD5) are coloured red. Those more common in the most deprived part of the population (SIMD1) are coloured blue.

**Table 8.7 Multivariate Analysis of Univariate Risk Factors (P<0.157) with Weights and sociodemographic variables removed**

| Factor | OR (95% CI) | P-value |
|---|---|---|
| Long term bowel condition | 0.46 (0.25 - 0.84) | 0.011 |
| Antacids | 0.41 (0.21 - 0.79) | 0.008 |
| Travel within Scotland | 2.25 (1.12 - 4.53) | 0.023 |
| Contact with Birds/Poultry | 0.27 (0.08 - 0.93) | 0.039 |
| Water source (public mains) | 6.27 (1.91 - 20.56) | 0.002 |
| Raw chicken washed before preparation | 0.26 (0.12 - 0.55) | <0.001 |
| Raw chicken cut up in the kitchen | 2.24 (1.15 - 4.37) | 0.018 |
| Eat beef oven-cooked, roasted or grilled | 1.74 (1.01 - 2.99) | 0.047 |
| Eat any unpasteurised dairy products (incl. milk and cheese) | 0.48 (0.23 - 0.99) | 0.047 |

Factors which were more common in the least deprived part of the population (SIMD5) are coloured red. Those more common in the most deprived part of the population (SIMD1) are coloured blue.

Note that P=0.05 gave the same results

**General details/Personal characteristics:** A number of the main personal characteristics were removed from the analysis as described above.

**Historical Health Conditions and Treatment:** Long term bowel condition and antacids were more commonly associated with cases from deprived areas (SIMD1) in all of the multivariate models. H2 blockers were also more commonly taken by deprived cases in the unweighted multivariate model (P<0.05).

**Travel History:** Travel within Scotland was associated with living in least deprived (SIMD5) areas for the P<0.157 weighted multivariate model only (Tables 8.7).

**Contact with animals:** Contact with birds and poultry was more common in cases from most deprived areas (SIMD1) for all of the multivariate models. The type of bird was predominantly a household pet (budgie (1), parrot (2), lovebirds (1) and aviary (1)) for the SIMD1 group. Whilst for those living in the least deprived areas (SIMD5) contact was predominantly with agricultural birds either at home or at a farm (chicken (6) and pheasant (1)). Wild birds are associated with *Campylobacter* infection in humans (Cody, McCarthy et al. 2015) and it is possible that caged birds could act as a potential reservoir though further research would be required to establish whether this was a significant risk.

**Exposure to water:** Having a public mains water source was associated with living in areas of least deprivation (SIMD5) in all of the models. This result is the same as for the models where the socioeconomic variables were included. It is worth noting that looking at the control questionnaires from chapter 7 that approximately 96% of those from both SIMD1 and SIMD5 areas had a public water supply. Hence, it appears that those cases living in areas of most deprivation (SIMD1) are less likely to have a PWS (86.7% Table 8.2) than SIMD5 cases and both SIMD1 and SIMD5 controls. Therefore it is unclear why being on a public water supply would be found to be a risk factor for those living in less deprived areas.

**Foods eaten – chicken:** Raw chicken washed before preparation was more common in cases living in the most deprived areas (SIMD1) for all of the multivariate models (see previous discussion on this factor). Raw chicken cut-up in the kitchen was more common in cases from least deprived areas (SIMD5) in the weighted multivariate models.

**Foods eaten - poultry other than chicken:** None were significant

**Foods Eaten – Other:** Eating beef oven cooked, roasted or grilled was more common whilst eating any unpasteurised dairy products was less common in least deprived areas in the weighted multivariable models only. Dietary consumption by socioeconomic group in Scotland has been investigated previously (www.food.gov.uk/sites/default/files/media/document/749-1-1324_Final_Report_2001-2009.pdf). This found that red meat consumption was greater in SIMD1 (most deprived) compared with SIMD5 (least deprived) which is the opposite pattern to that found here by consumption of beef in *Campylobacter* cases.

To the knowledge of the authors, performing a case-case analysis based on deprivation has not been conducted previously for campylobacteriosis and any other infectious disease. Hence, it is not really possible to discuss the results in terms of the previous literature. What is worth noting is that even after the socioeconomic variables have been removed there are few food related variables that come through in the multivariate regression. It may be that had further cases been available, leading to higher statistical power, then more factors may have achieved statistical significance.

## 8.3 Conclusions

The case-case analysis can highlight underlying differences in the populations it is comparing. Unsurprisingly but somewhat reassuringly, demonstrated that those socioeconomic factors that were indicators of higher deprivation (e.g. being on benefits) were significantly associated with cases living in the most deprived deprivation quintile (SIMD1). Whilst those factors associated with affluence or wealth (household income >£47k per year and living in a house >= 3 bedrooms) was significantly associated with cases living in the least deprived quintiles (SIMD5).

Despite the fairly low statistical power of the case-case analysis (see section 2.2.1.2) a number of factors were identified in the analysis. For all of the analysis, being on antacids, and for most of the analysis, taking H2 blockers, was significantly associated with cases from deprived areas (SIMD1). Taking PPIs were not associated with deprivation.

Being on a public mains source was associated with cases from the least deprived areas in all of the analysis conducted although it is unclear what this result means in the wider context as 96% of all cases from SIMD 1 and 5 have a public mains water source.

When the socioeconomic variables were removed from the analysis food and animal related risk factors became apparent. In particular the behaviour of washing raw chicken was more common in those cases from deprived areas whilst cutting raw chicken up in the kitchen was a more common behaviour in cases from the least deprived areas. Finally, contact with caged pet birds was more commonly associated with cases from the most deprived areas whilst bird contact in least deprived areas tended to be predominantly with chickens either at home or in a farm setting.

# 9. Conclusions for overall study

As a general overview a disease reporting pyramid was generated. This considered quantitation at three levels: community, GP (comprising patients presenting to GP, GP diagnosis and stool samples received at MMDL) and reported case. At the community level the likelihood of an individual with an episode of gastrointestinal disease making a GP appointment depended on the duration and severity of symptoms. This probability varied from practically 100% for those experiencing protracted bloody diarrhoea to only 1 in 48 for those with diarrhoea of short duration. At the GP level it was considered that GPs would generally be able to diagnose a GI infection as such. However only 1 in 8.9 of GP diagnoses would result in a stool sample being received by an MMDL as this is dependent both on the GP deciding that a stool sample should be taken as well as the individual then submitting such a sample. The step involving the MMDL submitting to the final reported case level was considered to be robust to lost cases because this only involves linkage between databases. The findings were in general agreement with those obtained from the IID2 study.

The study findings has been broken down into four parts:

- Part 1 summarises the analysis of reported cases and investigates whether the socioeconomic differences can be explained by the hypotheses given on page 5 of this report using evidence from the reported case study (Chapter 4), reporting biases (Chapter 3), case control (Chapter 7) and case-case chapters (Chapter 8).
- Part 2 summarises the analysis of hospitalised cases using the evidence from Chapter 4 and provides explanations of why there are differences by deprivation.
- Part 3 looks at the case-control study across the combined SIMD1 and SIMD5 populations and identifies campylobacteriosis risk factors.
- Part 4 summarises the challenges of conducting a case-control study in Scotland and includes learnings for future studies.

## Part 1. Reported Cases

## Why are there more cases in the least deprived Scottish population?

Poisson regression of reported cases suggests that deprivation is protective in both the univariate and multivariate analysis. The incidence of disease is 11.5% higher in rural than urban areas and it is known that deprived datazones are disproportionately urban.

Across all ages there tends to be a higher number of cases in the least deprived compared with most deprived quintiles except for young children. Further, campylobacteriosis incidence is highest in the older part of the population (>50

years). Again, this population segment includes a higher proportion of residents of least deprived areas.

## Identification of differences in risk factors between SIMD5 and SIMD1 cases (case-case study)

Unsurprisingly, socio-economic factors that were indicators of higher deprivation (e.g. being on benefits) were significantly associated with cases living in the most deprived quintile (SIMD1). Whilst those factors associated with affluence or wealth (household income >£47k per year and living in a house >= 3 bedrooms) were significantly associated with cases living in the least deprived quintiles (SIMD5).

## Those living in least deprived areas are more likely to report campylobacteriosis

Analysis of reported cases (Chapter 4) shows that there remains an excess of 19% of campylobacteriosis cases in the less deprived SIMD quintiles (i.e. SIMD2 to SIMD5). This is a 7% reduction from that observed in the 2000-2006 geography study but the figure remains statistically significant. Both univariate and multivariate Poisson regression of reported cases confirmed that as deprivation increases then campylobacteriosis incidence decreases (Chapter 4). The domestic (not foreign) case-control study (Chapter 7) found that cases from least deprived (SIMD5) areas were more likely to report campylobacteriosis than those from most deprived (SIMD1) areas but this was not statistically significant.

Five underlying factors contributing to those living in least deprived areas being more likely to report campylobacteriosis were investigated:

### 1) Difference in culinary habits

In particular, the behaviour of washing raw chicken was more common in those cases from deprived areas whilst cutting up raw chicken in the kitchen was a more common behaviour in cases from the least deprived areas (Chapter 8 case-case study).

### 2) Difference in levels of environmental exposure (water and animal exposures)

Being on a public mains source was associated with cases from the least deprived areas in all of the analyses conducted (once sociodemographic variables were removed; case-case study).

Contact with birds and poultry were more commonly associated with cases from the most deprived areas (once sociodemographic variables were removed; case-case study).

The only socioeconomic factor found to be associated with increased risk of campylobacteriosis was having a car <5 years old (domestic case-control study

Chapter 7). This is most likely a reflection of higher incidence rates in the more affluent who are more likely to own a newer car.

### 3) Difference in disease severity, hospitalisation or medication

The case-case study showed that taking antacids and, for most of the analysis, taking H2 blockers were significantly associated with cases from deprived areas (SIMD1). Taking PPIs was not associated with deprivation (Chapter 7).

The incidence of diagnoses by GPs of IID (RCG3) was higher in most deprived (SIMD1) compared with least deprived (SIMD5) areas (Chapter 4).

The likelihood of making a doctor's appointment did not vary between the least and most deprived populations based on number and duration of GI symptoms (Figure 3.2 and Figure 3.3) apart from prolonged "nausea or vomiting". This was more likely to lead to an individual from a least deprived background (SIMD5) making a doctor's appointment.

There was no significant difference between PTI and MMDL for GI reporting by deprivation (Table 3.5)

There was no difference in the duration (Fig 3.11) or frequency of symptoms for reported cases in the case-control study between SIMD1 and SIMD5 cases (Section 3.4.3). For hospitalised cases, there was also no difference in frequency or duration of symptoms by deprivation (Section 3.4.3)

### 4) Differences in reporting

Confronted with a patient with a potential GI infection most GPs stated that socio-economic factors were not considered important when considering whether to request a stool sample (section 3.3.1.5).

Looking at GP reporting of presumed GI infections (RCG3) there was a higher frequency among the most deprived (SIMD1) compared with SIMD5 areas. This is in keeping with the commonly found socio-economic gradient of disease incidence; but contrary to campylobacteriosis incidence which follows the opposite trend.

There is some evidence to suggest that people with campylobacteriosis living in SIMD1 (most deprived) areas within 1 km of a GP are less likely to attend than expected – though this is small (2% of SIMD1 cases; Section 4.4.4).

### 5) Difference due to foreign travel

Approximately 36% of the difference in reported cases between SIMD5 and SIMD1 can be explained by foreign travel.

GPs indicated that recent foreign travel was a very important consideration when requesting a stool sample and, in their view, for cases deciding to submit a stool sample. Since just over three times as many individuals from least (SIMD5) compared with most (SIMD1) deprived areas travel abroad then this will likely lead to more SIMD5 individuals having stool samples taken. However, from the

MMDL study, the overall number of stool samples taken appears lowest for SIMD5 though the overall sample size is small.

## Part 2. Hospitalised cases

For hospitalised cases, there is an opposite pattern to reported cases, with 9.2% excess in the two most deprived quintiles (SIMD1 and SIMD2).

This excess among the 40% of the population living in the most deprived areas can be attributed to two factors. The first being the high SIMD1 and SIMD2 populations close to hospital (<10 km) and secondly the high rate of hospitalisation within those populations. It is unclear why this rate is high but it is hypothesised that poorer general health and/or social circumstances may be contributing factors.

The incidence of campylobacteriosis hospitalisation has increased three-fold since 2005. This is mostly due to increases in the elderly (>65 years), whilst hospitalisation rates of children (both <5 years and the 5-14 year age groups) have been relatively stable throughout.

## Part 3. Risk Factors across the SIMD1 and SIMD5 populations

## Risk factors for domestic cases across SIMD1 and SIMD5 populations

The main food related risk factors were: eating chicken liver pâté prepared at home (PAF =5%); eating chicken lightly cooked (PAF =2 to 3%); eating chicken outside the home elsewhere (not restaurant, take-away or fast food) (PAF=8 to 9%); eating poultry (other than chicken) at a restaurant (PAF=5%) were all significant in all of the analysis. However, chicken consumption is a complex risk factor and can be "protective" depending on the setting and where it is prepared. For example raw chicken handled in the kitchen was (somewhat counter-intuitively) "protective".

The following non-food risk factors were also consistently significant in all of the analysis: being on PPIs (PAF 10 to 16%) and having white ethnicity (PAF 71 to 77%).

Using an indoor swimming pool/toddler pool was "protective" in all of the analysis. Contact with various animal groups, predominantly pets was mostly protective as was consumption of a number of other foods (e.g. pork and beef) but results were not consistent across all of the analysis

## Risk factors for foreign travel cases across SIMD1 and SIMD5 populations

Foreign travel was a significant risk factor in the case-control study. Two and a half times more cases than controls (24.4% compared with 9.4%) travelled abroad with an overnight stay in the 14 days before falling ill.

Increased risk of campylobacteriosis occurred when travelling to Asia including Turkey. The risk fell when travelling to North America.

The univariate analysis found two risk factors for foreign travel associated campylobacteriosis (eating chicken outside the home and eating lamb) but only the unweighted multivariate models revealed eating lamb to be a risk factor.

A number of risk factors were found to reduce the risk (e.g. household income >£47k per year, H2 blockers, dog contact and eating chicken prepared at home) but all of these were not consistently significant across the models.

The small numbers of questionnaires completed (146 cases and 52 controls) will have reduced the statistical power of the analysis.

## Part 4. The challenges of carrying out a case-control study across Scotland

To obtain all permissions (e.g. ethics and PBPP) took 17 months for this study. It is likely that future studies may be able to achieve this more quickly as for example the PBPP panel has now become established. However, in the authors view this is unlikely to require less than one year.

The case-control study is dependent on the goodwill of the NHS Health Boards to submit questionnaires. All were interested and supportive of the study but there were many competing priorities for their limited staff resource even though payment to cover their costs was made.

The response rates for the case-control study was low (22.7% for cases and 10.6% for controls). It may be that a shorter questionnaire, sending a reminder, and /or using another medium (telephone or in person interview) would increase the return rate. Compensation for the time taken to complete the questionnaire (e.g. a gift voucher) may also have increased the return rate. The availability of web-based electronic submission was not popular.

# 10. Implications for FSS

This is broken down into three areas. The first relates to how the research can be used to reduce campylobacteriosis. The second identifies methodological issues which should be addressed for future studies. The third identifies future research that should be considered.

**Utilising the research findings to reduce human campylobacteriosis**

- There continues to be an excess of reported cases of human campylobacteriosis in the least deprived population. There is not any strong evidence to show that this is due to reporting. As such it is important that this population is reminded of the causes of campylobacteriosis and what they can do to protect themselves.
- Undercooked (pink) chicken is a risk factor but has a fairly low population attributable fraction (explaining 2 to 3% of cases across most deprived (SIMD1) and least deprived (SIMD5) areas). This helps underpin the FSS pink chicken campaigns but suggests any resulting reductions in campylobacteriosis will be modest. However, ensuring chicken is properly cooked has the potential to reduce the risk of illness not only from *Campylobacter* but also from other pathogens that may contaminate chicken such as *Salmonella*.
- Eating chicken liver pâté prepared at home was a significant risk factor (PAF of 5%). This suggests that providing food safety advice to consumers and in recipes will be important to reduce this risk.
- Eating chicken outside the home elsewhere (at a friend or relative's house or a community/family gathering) was an important risk factor (PAF 8 to 9%). Providing food safety advice to consumers and in particular those who prepare the food at these events is warranted.
- Being on PPIs is an important risk factor and it may be valuable to target food safety advice to this population, particularly if they are elderly.
- Foreign travel is a risk factor for human campylobacteriosis particularly for those travelling to Asia including Turkey. Public health advice to these individuals has the potential to reduce the incidence of campylobacteriosis.
- Washing raw chicken was a more common behaviour in cases from the most deprived compared with cases from the least deprived populations (21.2% compared with 11.1%). Although washing chicken was not found to be a risk factor in the case-control analysis it may still be worth trying to communicate the message not to wash raw chicken to the most deprived (SIMD1) population.
- More cases of campylobacteriosis were associated with cases taking antacids (17.7% compared with 9.9%) or H2 blockers (5.3% compared with 1.8%) in the most deprived compared with the least deprived population. It may be worth communicating this risk to these groups so

that they take additional precautions to reduce the risk of campylobacteriosis.

**Methodological considerations for future studies:**

- Case-control studies are an accepted methodology to identify the putative sources of infectious diseases whether that be from an outbreak or from sporadic cases. The ethics required to progress these studies is important but challenging. Efforts should be made to streamline this process.
- Future case control studies should consider means of incentivising both cases and controls to participate in such a study and to evaluate both the mechanism and format of questionnaire to increase the likelihood of response.
- Whole genome sequencing is now becoming commonplace in the study of infectious diseases and using this technique for both source tracking and source attribution is now well established. In future case control studies should where practical include the whole genome sequencing of a representative number of isolates. The combination of the two methods provides additional evidence in elucidating the source of human disease.

**Future Research:**

- Cases of campylobacteriosis associated with foreign travel may be more likely to be reported than infections acquired domestically. Further research should be conducted to determine whether this is the case. The findings are likely to be relevant to other gastrointestinal infections such as salmonellosis.
- The high incidence of discharges in the SIMD1 and SIMD2 populations within 10 km of a hospital compared with the less deprived quintiles should be investigated to see if this is due to differences in general health of the population or some other factor.
- The rapidly increasing rate of hospitalisations in the >65 year old population warrants further investigation. This trend is increasing and with the Scottish population ageing it will become more important in future years.

# Acknowledgements

The authors would like to acknowledge the help and input from the Health Protection Teams from each of the health boards who have participated in this study. It would not have been possible to complete this study without their help. Health Protection Scotland and in particular Alison Smith-Palmer, is acknowledged for providing the data on reported cases from across Scotland and advice during the study. Colleagues at the Grampian Data Safe Haven (DaSH) are acknowledged for their help enabling the use of this facility for carrying out the project work as well as Patricia Burns and colleagues at Research Governance at University of Aberdeen/ NHS Grampian. Carole Morris at eDRIS is acknowledged for her help and advice with the PBPP application as well as coordinating hospitalisation and reported case data uploads to DaSH. There have also been a number of scientists who have provided helpful advice during the study and this includes: Steen Ethelberg and Katrin Kuhn from Statens Serum Institute in Denmark, Lapo-Mughini-Gras from RIVM in the Netherlands and Professor Noel McCarthy from the University of Warwick.

# Recent outputs from this and related studies

**The following talks, presentations and posters**:

Rotariu Ovidiu, Forbes Ken, McGuigan Chris* and Strachan Norval (2017) A study elucidating the socio-demographics of *Campylobacter* infection in Scotland. CHRO, 10-14 September 2017, Nantes, France.

Strachan Norval, Rotariu Ovidiu, Macrae Marion, Lopes Bruno, Ramjee Meenakshi and Forbes Ken (2017) Whole genome sequencing and empirical epidemiology identifies disparate aetiologies between *Campylobacter jejuni* sequence types ST50 and ST61. CHRO, 10-14 September 2017, Nantes, France.

Strachan NJC (2016) A pot-pourri of GI pathogen anecdotes: "*Campylobacter*, Anisakiasis and STEC". University of Aberdeen, Roslin, Edinburgh

Stephen P Rushton, Roy A Sanderson, Peter J Diggle, Mark DF Shirley, Alasdair P Blain, Iain Lake, James A Maas, William DK Reid, Jo Hardstaff, Nicola Williams, Natalia R Jones, Daniel Rigby, Norval JC Strachan, Ken J Forbes, Paul R Hunter, Thomas J Humphrey, Sarah J O'Brien (2019) Climate, human behaviour or environment: individual-based modelling of *Campylobacter* seasonality and strategies to reduce disease burden. Journal of Translational Medicine, 17 (1), 34.

Natalia R Jones, Caroline Millman, Mike van der Es, Miroslava Hukelova, Ken J Forbes, Catherine Glover, Sam Haldenby, Paul R Hunter, Kathryn Jackson, Sarah J O'Brien, Dan Rigby, Norval JC Strachan, Nicola Williams, Iain R Lake, Enigma consortium. Novel sampling method for assessing human-pathogen interactions in the natural environment using boot socks and citizen scientists, with application to *Campylobacter* seasonality. Appl. Environ. Microbiol. 83 (14), e00162-17

**Research Projects which have been assisted by this project**:

FSS (2015-2017) i-CaMPS-4. Employing source attribution and molecular epidemiology to measure the impact of interventions on human campylobacteriosis in Scotland, £230k.

Forbes K, Strachan N (2014-17) "*Campylobacter* disease in Nigeria" £70,000 from University of Aberdeen Elphinstone PhD Scholarship.

O'Brien SJ, Bennett M, Diggle PJ, Forbes KJ, Griffith R, Humphrey T, Hunter P, Lake I, Rigby D, Rushton S, Strachan NJC, Wadsworth R, Winstanley C, Wren B (2012 -2017) "Sources, Seasonality, Transmission and Control: *Campylobacter* and human behaviour in a changing environment." £3,419,121 from MRC.

Forbes K, Strachan N, Stevens M, Psifidi A, Vervelde L (2016-19) "A systems-wide approach to the control of *Campylobacter* in the food chain: exploiting genetic variation." £806,000 from Scottish Government RESAS.

# References


ACMSF, 2018. *Epidemiology of Foodborne Infections Group Report.*

ADAK, G.K., LONG, S.M. and O'BRIEN, S.J., 2002. Trends in indigenous foodborne disease and deaths, England and Wales: 1992 to 2000. *Gut,* **51**(6), pp. 832-841.

ANON, 2016. Employing Source Attribution and Molecular Epidemiology to measure the impact of interventions on human campylobacteriosis in Scotland. *Food Standards Agency- Scotland Contract S14054,* , pp. 1-89.

ANON, 2012. *Scottish Index of Multiple Deprivation (SIMD) 2012.* Edinburgh, Scotland, UK: The Scottish Government.

ANON, 2010. *Private water supplies as a risk factor for Campylobacter infection in Aberdeen city and Aberdeenshire.* S14023.

ANON., 2017. *i-CaMPS4 Employing Source Attribution and Molecular Epidemiology to measure the impact of interventions on human campylobacteriosis in Scotland. An extension focused on the role of Scottish broiler production on human campylobacteriosis cases.*

ANON., 2007. *Factors Associated with geographical and temporal variation in Campylobacteriosis in humans.*

BECH, M. and LAURIDSEN, J., 2009. Exploring spatial patterns in general practice expenditure. *The European journal of health economics : HEPAC : health economics in prevention and care,* **10**(3), pp. 243-254.

BELONGIA, E.A., CHYOU, P.H., GREENLEE, R.T., PEREZ-PEREZ, G., BIBB, W.F. and DEVRIES, E.O., 2003. Diarrhea incidence and farm-related risk factors for Escherichia coli O157:H7 and Campylobacter jejuni antibodies among rural children. *The Journal of infectious diseases,* **187**(9), pp. 1460-1468.

BEMIS, K., MARCUS, R. and HADLER, J.L., 2014. Socioeconomic status and campylobacteriosis, Connecticut, USA, 1999-2009. *Emerging infectious diseases,* **20**(7), pp. 1240-1242.

BESSELL, P.R., MATTHEWS, L., SMITH-PALMER, A., ROTARIU, O., STRACHAN, N.J., FORBES, K.J., COWDEN, J.M., REID, S.W. and INNOCENT, G.T., 2010. Geographic determinants of reported human Campylobacter infections in Scotland. *BMC public health,* **10**, pp. 423.

BLASER, M.J., 1997. Epidemiologic and clinical features of *Campylobacter jejuni* infections. *The Journal of infectious diseases,* **176 Suppl 2**, pp. S103-105.

BURSAC, Z., GAUSS, C.H., WILLIAMS, D.K. and HOSMER, D.W., 2008. Purposeful selection of variables in logistic regression. *Source code for biology and medicine,* **3**, pp. 17-0473-3-17.

CAULCUTT, R., 1983. *Statistics in Research and Development.* London, UK.: Chapman and Hall Ltd.

CLIFFORD-BLAIR, R. and HIGGINS, J.J., 1980. A Comparison of the Power of Wilcoxon's Rank-Sum Statistic to That of Student's t Statistic under Various Nonnormal Distributions. *Journal of Educational Statistics,* **5**, pp. 309-335.

CODY, A.J., MCCARTHY, N.D., BRAY, J.E., WIMALARATHNA, H.M.L., COLLES, F.M., VAN RENSBURG, M.J.J., DINGLE, K.E., WALDENSTROM, J. and MAIDEN, M.C.J., 2015. Wild bird-associated Campylobacter jejuni isolates are a consistent source of human disease, in Oxfordshire, United Kingdom. *Environmental Microbiology Reports,* **7**(5), pp. 782-788.

COX, D.R., 1958. The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society. Series B (Methodological),* **20**, pp. 215-242.

CROASMUN, J.T. and OSTROM, L., 2011. Likert-Type Scales in the Social Sciences, Journal of Adult Education. *Journal of Adult Education,* **40(1)**, pp. 19-22.

DANIS, K., DI RENZI, M., O'NEILL, W., SMYTH, B., MCKEOWN, P., FOLEY, B., TOHANI, V. and DEVINE, M., 2009. Risk factors for sporadic Campylobacter infection: an all-Ireland case-control study. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin,* **14**(7),.

DEFRA, 2019. *United Kingdom Poultry and Poultry Meat Statistics – January 2019.*

DOMINGUES, A.R., PIRES, S.M., HALASA, T. and HALD, T., 2012. Source attribution of human campylobacteriosis using a meta-analysis of case-control studies of sporadic infections. *Epidemiology and infection,* **140**(6), pp. 970-981.

DOORDUYN, Y., VAN DEN BRANDHOF, W.E., VAN DUYNHOVEN, Y.T.H.P., BREUKINK, B.J., WAGENAAR, J.A. and VAN PELT, W., 2010. Risk factors for indigenous Campylobacter jejuni and Campylobacter coli infections in The Netherlands: a case-control study. *Epidemiology and infection,* **138**(10), pp. 1391-1404.

EFIRD, J.T., 2013. Computing power and sample size for informational odds ratio. *International journal of environmental research and public health,* **10**(10), pp. 5239-5243.

EKDAHL, K. and ANDERSSON, Y., 2004. Regional risks and seasonality in travel-associated campylobacteriosis. *BMC infectious diseases [computer file],* **4**(1), pp. 54.

ETHELBERG, S., SIMONSEN, J., GERNER-SMIDT, P., OLSEN, K.E. and MOLBAK, K., 2005. Spatial distribution and registry-based case-control analysis of Campylobacter infections in Denmark, 1991-2001. *American Journal of Epidemiology,* **162**(10), pp. 1008-1015.

FAJO-PASCUAL, M., GODOY, P., FERRERO-CANCER, M. and WYMORE, K., 2010. Case-control study of risk factors for sporadic Campylobacter infections in northeastern Spain. *European journal of public health,* **20**(4), pp. 443-448.

FEODOROFF, B., LAUHIO, A., ELLSTROM, P. and RAUTELIN, H., 2011. A nationwide study of Campylobacter jejuni and Campylobacter coli bacteremia in Finland over a 10-year period, 1998-2007, with special reference to clinical characteristics and antimicrobial susceptibility. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America,* **53**(8), pp. e99-e106.

FISHER, R.A., 1935. The logic of inductive inference. *Journal of the Royal Statistical Society,* **1**, pp. 39-82.

FORBES, K.J., GORMLEY, F.J., DALLAS, J.F., LABOVITIADI, O., MACRAE, M., OWEN, R.J., RICHARDSON, J., STRACHAN, N.J., COWDEN, J.M., OGDEN, I.D. and MCGUIGAN, C.C., 2009. Campylobacter immunity and coinfection following a large outbreak in a farming community. *Journal of clinical microbiology,* **47**(1), pp. 111-116.

GALANIS, E., MAK, S., OTTERSTATTER, M., TAYLOR, M., ZUBEL, M., TAKARO, T.K., KUO, M. and MICHEL, P., 2014. The association between campylobacteriosis, agriculture and drinking water: a case-case study in a region of British Columbia, Canada, 2005-2009. *Epidemiology and infection,* **142**(10), pp. 2075-2084.

GARDNER, W., MULVEY, E.P. and SHAW, E.C., 1995. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychol. Bull.,* **118**, pp. 392-404.

GIESECKE, J., 2002. *Modern infectious disease epidemiology.* 2nd Edition edn. London, UK: Arnold.

GILLESPIE, I.A., O'BRIEN, S.J., FROST, J.A., ADAK, G.K., HORBY, P., SWAN, A.V., PAINTER, M.J., NEAL, K.R. and CAMPYLOBACTER SENTINEL SURVEILLANCE SCHEME COLLABORATORS, 2002. A case-case comparison of Campylobacter coli and Campylobacter jejuni infection: a tool for generating hypotheses. *Emerging infectious diseases,* **8**(9), pp. 937-942.

GILLESPIE, I.A., O'BRIEN, S.J., PENMAN, C., TOMPKINS, D., COWDEN, J. and HUMPHREY, T.J., 2008. Demographic determinants for Campylobacter infection in England and Wales: implications for future epidemiological studies. *Epidemiology and infection,* **136**(12), pp. 1717-1725.

GILLESPIE, I.A., O'BRIEN, S.J. and BOLTON, F.J., 2009. Age Patterns of Persons with Campylobacteriosis, England and Wales, 1990-2007. *Emerging Infectious Diseases,* **15**(12), pp. 2046-2048.

GORMLEY, F.J., MACRAE, M., FORBES, K.J., OGDEN, I.D., DALLAS, J.F. and STRACHAN, N.J., 2008. Has retail chicken played a role in the decline of human campylobacteriosis? *Applied and Environmental Microbiology,* **74**(2), pp. 383-390.

GREENLAND, S. and ROBINS, J., 1988. Conceptual Problems in the Definition and Interpretation of Attributable Fractions. *American Journal of Epidemiology,* **128**(6), pp. 1185-1197.

GREGORY, P.M., MALKA, E.S., KOSTIS, J.B., WILSON, A.C., ARORA, J.K. and RHOADS, G.G., 2000. Impact of geographic proximity to cardiac revascularization services on service utilization. *Medical care,* **38**(1), pp. 45-57.

HANEUSE, S.J. and WAKEFIELD, J.C., 2007. Hierarchical models for combining ecological and case-control data. *Biometrics,* **63**(1), pp. 128-136.

HAVELAAR, A.H., VAN PELT, W., ANG, C.W., WAGENAAR, J.A., VAN PUTTEN, J.P.M., GROSS, U. and NEWELL, D.G., 2009. Immunity to Campylobacter: its role in risk assessment and epidemiology. *Critical reviews in microbiology,* **35**(1), pp. 1-22.

HOSMER, D.W., LEMESHOW, S. and STURDIVANT, R.X., 2013. *Applied Logistic Regression.* 3rd edn. Hoboken, New Jersey, USA: Wiley & Sons.

HOWIE, H., MUKERJEE, A., COWDEN, J., LEITH, J. and REID, T., 2003. Investigation of an outbreak of Escherichia coli O157 infection caused by environmental exposure at a scout camp. *Epidemiology and infection,* **131**(3), pp. 1063-1069.

JONES, A.K., RIGBY, D., BURTON, M., MILLMAN, C., WILLIAMS, N.J., JONES, T.R., WIGLEY, P., O'BRIEN, S.J., CROSS, P. and ENIGMA CONSORTIUM, 2016. Restaurant Cooking Trends and Increased Risk for Campylobacter Infection. *Emerging infectious diseases,* **22**(7), pp. 1208-1215.

KLEINBAUM, D.G. and KLEIN, M., 2010. *Logistic Regression A Self-Learning Text.* 3rd edn. New York, USA: Springer.

KUHN, K.G., NIELSEN, E.M., MOLBAK, K. and ETHELBERG, S., 2018. Determinants of sporadic *Campylobacter* infections in Denmark: a nationwide

case-control study among children and young adults. *Clinical Epidemiology,* **10**, pp. 1695-1707.

LIN, G., ALLAN, D.E. and PENNING, M.J., 2002. Examining distance effects on hospitalizations using GIS: a study of three health regions in British Columbia, Canada. *Canada Environment and Planning,* **34**, pp. 2037-2053.

LITTLE, C., GORMLEY, F., RAWAL, N. and RICHARDSON, J., 2010. A recipe for disaster: outbreaks of campylobacteriosis associated with poultry liver pate in England and Wales. *Epidemiology and infection,* **138**(12), pp. 1691-1694.

MACRITCHIE, L., HUNTER, C. and STRACHAN, N.J.C., 2013. A population based exposure assessment of risk factors associated with gastrointestinal pathogens: a *Campylobacter* study. *Epidemiology and Infection,* **141**, pp. 97-986.

MANLY, B.F.J., 2007. *Randomization, Bootsrap and Monte Carlo Methods in Biology.* 3rd Edition edn. Boca Raton, Florida, USA: Chapman & Hall/CRC.

MANN, H.B. and WHITNEY, D.R., 1947. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics,* **18(1)**, pp. 50-60.

MCCARTHY, N. and GIESECKE, J., 1999. Case-case comparisons to study causation of common infectious diseases. *International journal of epidemiology,* **28**(4), pp. 764-768.

MCLOONE, P., 2004. *Carstairs scores for Scottish Postcode Sectors from the 2001 census.* Glasgow, UK.: MRC Social and Public Health Sciences Unit.

MIETTINEN, O.S., 1974. Proportion of a disease caused or prevented by a given exposure, trait or intervention. *Am J Epidemiol,* **99**, pp. 325-332.

MOY, P. and MURPHY, J., 2016. Problems and Prospects in Survey Research. *Journalism & Mass Communication Quarterly,* **93**(1), pp. 16-37.

MUGHINI-GRAS, L., SMID, J.H., WAGENAAR, J.A., DE BOER, A., HAVELAAR, A.H., FRIESEMA, I.H., FRENCH, N.P., GRAZIANI, C., BUSANI, L. and VAN PELT, W., 2014. Campylobacteriosis in returning travellers and potential secondary transmission of exotic strains. *Epidemiology and infection,* **142**(6), pp. 1277-1288.

MURPHY, H.M., THOMAS, M.K., SCHMIDT, P.J., MEDEIROS, D.T., MCFAYDEN, S. and PINTAR, K.D.M., 2016.
Estimating the burden of acute gastrointestinal illness due to Giardia, Cryptosporidium, Campylobacter, E. coli O157 and norovirus associated with private wells and small water systems in Canada
. *Epidemiology and infection,* **144**, pp. 1355-1370.

NICHOLS, G., LANE, C., ASGARI, N., VERLANDER, N.Q. and CHARLETT, A., 2009. Rainfall and outbreaks of drinking water related disease and in England and Wales. *Journal of water and health,* **7**(1), pp. 1-8.

NICHOLS, G.L., RICHARDSON, J.F., SHEPPARD, S.K., LANE, C. and SARRAN, C., 2012. Campylobacter epidemiology: a descriptive study reviewing 1 million cases in England and Wales between 1989 and 2011. *BMJ open,* **2**(4),.

OGDEN, I.D., DALLAS, J.F., MACRAE, M., ROTARIU, O., REAY, K.W., THOMSON, A.P., SHEPPARD, S.K., MAIDEN, M.C., FORBES, K.J. and STRACHAN, N.J.C., 2009. *Campylobacter* excreted into the environment by animal sources: prevalence, concentration shed and host association. *Environmental microbiology,* **6**(10), pp. 1161-1170.

OLOWOKURE, B., HAWKER, J., WEINBERG, J., GILL, N. and SUFI, F., 1999. Deprivation and hospital admission for infectious intestinal diseases. *Lancet,* **353**(9155), pp. 807-808.

OSGOOD, D.W., 2000. Poisson-Based Regression Analysis of Aggregate Crime Rates. *Journal of Quantitative Criminology,* **16**, pp. 21-43.

PERRY, J.L., DEMPSTER, M. and MCKAY, M.T., 2017. Academic Self-Efficacy Partially Mediates the Relationship between Scottish Index of Multiple Deprivation and Composite Attainment Score. *Frontiers in Psychology,* **8**, pp. 1899.

POGREBA-BROWN, K. and BARRETT, E., 2018. Campylobacter and Ethnicity-A Case-Case Analysis to Determine Differences in Disease Presentation and Risk Factors. *Foodborne pathogens and disease,* **15**(5), pp. 277-284.

ROSENBERG GOLDSTEIN, R.E., CRUZ-CANO, R., JIANG, C., PALMER, A., BLYTHE, D., RYAN, P., HOGAN, B., WHITE, B., DUNN, J.R., LIBBY, T., TOBIN-D'ANGELO, M., HUANG, J.Y., MCGUIRE, S., SCHERZINGER, K., LEE, M.L. and SAPKOTA, A.R., 2016. Association between community socioeconomic factors, animal feeding operations, and campylobacteriosis incidence rates: Foodborne Diseases Active Surveillance Network (FoodNet), 2004-2010. *BMC infectious diseases,* **16**, pp. 354-016-1686-9.

ROSNER, B.M., SCHIELKE, A., DIDELOT, X., KOPS, F., BREIDENBACH, J., WILLRICH, N., GOLZ, G., ALTER, T., STINGL, K., JOSENHANS, C., SUERBAUM, S. and STARK, K., 2017. A combined case-control and molecular source attribution study of human Campylobacter infections in Germany, 2011-2014. *Scientific reports,* **7**(1), pp. 5139-017-05227-x.

ROUX, F., SPROSTON, E.L., ROTARIU, O., MACRAE, M., SHEPPARD, S.K., BESSELL, P.R., SMITH-PALMER, A., COWDEN, J.M., MAIDEN, M.C., FORBES, K.J. and STRACHAN, N.J.C., 2013. Elucidating the aetiology of human *Campylobacter coli* infections. *PLoS One,* **8**(5), pp. e64504.

SCALLAN, E., GRIFFIN, P.M., MCLEAN, H.Q. and MAHON, B.E., 2018. Hospitalisations due to bacterial gastroenteritis: A comparison of surveillance and hospital discharge data. *Epidemiology and infection,* **146**(8), pp. 954-960.

SEARS, A., 2009. *Campylobacteriosis in New Zealand: describing the marked decline in notifications 2007-2008 and discussing its possible causes.*, University of Otago.

SHEPPARD, S.K., DALLAS, J.F., MACRAE, M., MCCARTHY, N.D., SPROSTON, E.L., GORMLEY, F.J., STRACHAN, N.J., OGDEN, I.D., MAIDEN, M.C. and KEN, J.F., 2009. Campylobacter genotypes from food animals, environmental sources and clinical disease in Scotland 2005/6. *International journal of food microbiology,* .

SIMMONS, K., GAMBHIR, M., LEON, J. and LOPMAN, B., 2013. Duration of immunity to norovirus gastroenteritis. *Emerging infectious diseases,* **19**(8), pp. 1260-1267.

SIMONSEN, J., FRISCH, M. and ETHELBERG, S., 2008. Socioeconomic risk factors for bacterial gastrointestinal infections. *Epidemiology,* **19**(2), pp. 282-290.

SMITH-PALMER, A. and COWDEN, J., 2010. *Private water supplies as a risk factor for Campylobacter infection in Aberdeen City and Aberdeenshire.*

SPENCER, S.E.F., MARSHALL, J., PIRIE, R., CAMPBELL, D., BAKER, M.G. and FRENCH, N.P., 2012. The spatial and temporal determinants of campylobacteriosis notifications in New Zealand, 2001-2007. *Epidemiology and infection,* **140**(9), pp. 1663-1677.

STRACHAN, N.J.C., GORMLEY, F.J., ROTARIU, O., OGDEN, I.D., MILLER, G., DUNN, G.M., SHEPPARD, S.K., DALLAS, J.F., REID, T.M.S., HOWIE, H., MAIDEN, M.C. and FORBES, K.J., 2009. Attribution of *Campylobacter* infections in northeast Scotland to specific sources using multi-locus sequence typing (MLST). *The Journal of infectious diseases,* **199**(15 April), pp. 1205-1208.

STRACHAN, N.J., GORMLEY, F.J., ROTARIU, O., OGDEN, I.D., MILLER, G., DUNN, G.M., SHEPPARD, S.K., DALLAS, J.F., REID, T.M., HOWIE, H., MAIDEN, M.C. and FORBES, K.J., 2009. Attribution of campylobacter infections in northeast Scotland to specific sources by use of multilocus sequence typing. *The Journal of infectious diseases,* **199**(8), pp. 1205-1208.

STRACHAN, N.J., WATSON, R.O., NOVIK, V., HOFREUTER, D., OGDEN, I.D. and GALAN, J.E., 2008. Sexual dimorphism in campylobacteriosis. *Epidemiology and infection,* **136**(11), pp. 1492-1495.

STRACHAN, N.J.C., MACRAE, M., THOMSON, A., ROTARIU, O., OGDEN, I.D. and FORBES, K.J., 2012. Source attribution, prevalence and enumeration of Campylobacter spp. from retail liver. *International journal of food microbiology,* **153**(1-2), pp. 234-236.

STRACHAN, N.J.C., ROTARIU, O., MACRAE, M., SHEPPARD, S.K., SMITH-PALMER, A., COWDEN, J., MAIDEN, M.C.J. and FORBES, K.J., 2013. Operationalising Factors That Explain the Emergence of Infectious Diseases: A Case Study of the Human Campylobacteriosis Epidemic. *Plos One,* **8**(11), pp. e79331.

TAM, C., VIVIANI, L., ADAK, B., BOLTON, E., DODDS, J., COWDEN, J., EVANS, M., GRAY, J., HUNTER, P., JACKSON, K., LETLEY, L., NEAL, K., RAIT, G., SMITH, G., SMYTH, B., TOMPKINS, D., VAN DER ES, M., RODRIGUES, L. and O'BRIEN, S., 2011. *The second study of infectious intestinal disease in the community (IID2).* B18021.

TAM, C.C., RODRIGUES, L.C., PETERSEN, I., ISLAM, A., HAYWARD, A. and O'BRIEN, S.J., 2006. Incidence of Guillain-Barre syndrome among patients with Campylobacter infection: a general practice research database study. *The Journal of infectious diseases,* **194**(1), pp. 95-97.

TAM, C.C., RODRIGUES, L.C., VIVIANI, L., DODDS, J.P., EVANS, M.R., HUNTER, P.R., GRAY, J.J., LETLEY, L.H., RAIT, G., TOMPKINS, D.S., O'BRIEN, S.J. and IID2 STUDY EXECUTIVE COMM, 2012. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut,* **61**(1),.

TEUNIS, P.F., FALKENHORST, G., ANG, C.W., STRID, M.A., DE VALK, H., SADKOWSKA-TODYS, M., ZOTA, L., KUUSI, M., ROTA, M.C., SIMONSEN, J.B., MOLBAK, K., VAN DUYNHOVEN, Y.T. and VAN PELT, W., 2013. Campylobacter seroconversion rates in selected countries in the European Union. *Epidemiology and infection,* **141**(10), pp. 2051-2057.

TUKEY, J.W., 1949. Comparing Individual Means in the Analysis of Variance. *Biometrics,* **5**, pp. 99-114.

VARGA, C., MIDDLETON, D., WALTON, R., SAVAGE, R., TIGHE, M., ALLEN, V., AHMED, R. and ROSELLA, L., 2012. Evaluating risk factors for endemic human Salmonella Enteritidis infections with different phage types in Ontario, Canada using multinomial logistic regression and a case-case study approach. *Bmc Public Health,* **12**, pp. 866.

WAGENAAR, J.A., FRENCH, N.P. and HAVELAAR, A.H., 2013. Preventing Campylobacter at the Source: Why Is It So Difficult? *Clinical Infectious Diseases,* **57**(11), pp. 1600-1606.

ZIA, S., WAREING, D., SUTTON, C., BOLTON, E., MITCHELL, D. and GOODACRE, J.A., 2003. Health problems following Campylobacter jejuni enteritis in a Lancashire population. *Rheumatology (Oxford, England),* **42**(9), pp. 1083-1088.